



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Eđitim Bilimleri Anabilim Dalı
Eđitimde Ölçme ve Deđerlendirme Bilim Dalı

ANGOFF, YES/NO VE EBEL STANDART BELİRLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Ceylan Gündeđer

Yüksek Lisans Tezi

Ankara, 2012

ANGOFF, YES/NO VE EBEL STANDART BELİRLEME YÖNTEMLERİNİN
KARŞILAŞTIRILMASI

Ceylan Gündeğer

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Eğitim Bilimleri Anabilim Dalı

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Yüksek Lisans Tezi

Ankara, 2012

KABUL VE ONAY

Ceylan Gündeğer tarafından hazırlanan "Angoff, Yes/No ve Ebel Standart Belirleme Yöntemlerinin Karşılaştırılması" başlıklı bu çalışma, 13.01.2012 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Yüksek Lisans tezi olarak kabul edilmiştir.



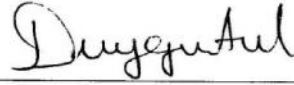
Prof. Dr. Selahattin GELBAL (Başkan)




Doç. Dr. Hülya KELECİOĞLU



Doç. Dr. Nuri DOĞAN (Danışman)



Doç. Dr. Duygu ANIL



Yrd. Doç. Dr. Neşe GÜLER

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Şahap Armağan TARIM

Enstitü Müdürü

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kağıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin/Raporumun tamamı her yerden erişime açılabilir.
- Tezim/Raporum sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin/Raporumun 1. yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

13.01.2012

Ceylan Gündeğer



Eđitimci ve zair dedem Tefik Gündeđer'e ...

TEŞEKKÜR

Öncelikle akademik hayata atılmamda benden hiçbir desteğini esirgememiş olan sayın hocam Prof. Dr. Doğan Nadi LEBLEBİCİ'ye teşekkürlerimi sunarım.

Yüksek lisans ders dönemimde bilgi derinliğine ve öğretme sanatına hayran olduğum, tez aşamamda ise kişiliği ile de bana örnek olmuş olan, her problemimde neşeli tavırlarıyla ve bana olan güveniyle anında “*jetonumu düşüren*” danışmanım Doç. Dr. Nuri DOĞAN'a çok teşekkür ederim.

Yüksek lisans eğitimim boyunca önemli katkılarda bulunan ve cadılarından biri olmaktan gurur duyduğum değerli hocam Prof. Dr. Selahattin GELBAL'a; beni tez konumla tanıştıran ve bu alanda çalışma yapmamda yüreklendiren sayın hocam Doç. Dr. Hülya KELECİOĞLU'na; ders dönemimde kendisinden çok şey öğrendiğim ve dersi kapsamında bu çalışmada kullandığım veri toplama aracını geliştirmeye başladığım sayın hocam Doç. Dr. Duygu ANIL'a ve lisans dönemimde ders hocam olup bu alana ilgi duymamda temel etken olan, önemli katkılarıyla jürimde bulunmasından mutluluk duyduğum hocam Yrd. Doç. Dr. Neşe GÜLER'e ne kadar teşekkür etsem azdır ...

Dünyada sahip olduğum tek “*gerçek*” şey olduklarına inandığım aileme, beni her konuda destekledikleri ve bana güvendikleri için teşekkür ederim. Özellikle veri toplama aşamasında benimle oradan oraya koşturmuş ve belki de benden daha fazla yorulmuş olan canım Babam'a çok teşekkür ederim. Ayrıca ailemizin bir ferdi olan ve tüm stresli anlarımda mırıldamasıyla beni rahatlatan kedim Burhan'a da teşekkürü bir borç bilirim.

Zor zamanlarımda benden desteklerini esirgememiş olan başta değerli dostum Arş. Gör. Özlem Altındağ'a ve babasına, yazar arkadaşım Seran Demiral'a, şu an ismini saymadığım yurt içi ve yurt dışındaki tüm arkadaşlarıma, akrabalarıma, hocalarıma ve özellikle de çalışma kapsamında yer almaktan geri durmayan öğretmenlerime sonsuz teşekkürler ...

ÖZET

GÜNDEĞER, Ceylan. Angoff, Yes/No ve Ebel *Standart Belirleme Yöntemlerinin Karşılaştırılması*, Yüksek Lisans Tezi, Ankara, 2012.

Bu araştırmada Angoff, Yes/No ve Ebel standart belirleme yöntemleri ile elde edilen kararlar ve kesme puanları karşılaştırılmıştır.

Araştırmanın öğrenci verilerini, ilköğretim 4. sınıf matematik dersi “Bölme ve Kesirler” alt öğrenme alanlarına yönelik araştırmacı tarafından hazırlanmış 20 soruluk bir başarı testinden alınan 489 öğrenci puanı oluşturmuştur. Araştırmada kullanılan diğer veriler ise 17 uzmanın hazırlanan bu başarı testindeki maddeler için üç farklı standart belirleme yöntemine göre vermiş oldukları kararlardır.

Verilerin analizinde üç farklı standart belirleme yöntemine göre başarılı sayılan öğrenci yüzdeleri arasındaki farklılıklar bağımlı iki oran arasındaki farkın testi ile sınanmıştır. Bu üç yöntemden elde edilen kesme puanlarına göre öğrenci puanları başarılı-başarısız olmak üzere yapay ikili hâle getirilip aralarındaki uyuma Cohen’in Kappa istatistiği ile bakılmıştır. Üç yönteme ait kesme puanları arasındaki uyuma Pearson Momentler Çarpımı Korelasyon Katsayısı ile ve kesme puanlarının ortalamaları arasındaki farka da bağımlı gruplar t testi ile bakılmıştır. Yöntemlerdeki uzmanlar arası uyumun belirlenmesi amacıyla Angoff ve Ebel yöntemleri için Kendall’ın uyum katsayısı (W); Yes/No yöntemi için ise Cohen’in Kappa istatistiği hesaplanmıştır. Ayrıca yöntemlere ait kararların varyans bileşenleri ve varyans yüzdeleri genellenebilirlik kuramı ile incelenmiş; her bir yöntem için en uygun puanlayıcı sayısı da D çalışması ile belirlenmiştir.

Bu araştırma sonunda bu üç farklı standart belirleme yöntemine göre başarılı sayılan öğrenci yüzdeleri arasında 0,01 hata düzeyinde manidar farklılık olduğu görülmüştür. Bu üç yöntemden elde edilen kesme puanlarına göre öğrenci puanları başarılı-başarısız olmak üzere yapay ikili hâle getirildiğinde Angoff ve Yes/No yöntemleri ile Angoff ve Ebel yöntemleri arasında 0,01 hata düzeyinde yüksek bir uyum; Yes/No ve Ebel yöntemleri arasında ise 0,01 hata düzeyinde makul bir uyum olduğu sonucuna ulaşılmıştır. Üç yönteme ait kesme puanları bakımından Angoff ve Yes/No yöntemleri ile belirlenen minimum geçme puanları (MGP) arasında pozitif ve orta düzeyde bir ilişki olduğunu ve iki yönteme ait MGP’lerin ortalamaları arasında da manidar bir farklılığın olmadığını; Angoff ve Ebel yöntemleri ile belirlenen MGP’ler arasında

pozitif ve yüksek düzeyde bir ilişki olduğu ve iki yonteme ait MGP'lerin ortalamaları arasında ise manidar bir farklılığın olduğunu; Yes/No ve Ebel yöntemleri ile belirlenen MGP'ler arasında ise anlamlı bir ilişkinin olmadığı ve iki yonteme ait MGP'lerin ortalamaları arasında ise manidar bir farklılığın olduğunu söyleyebiliriz. Yöntemlerdeki uzmanlar arası uyumun ise makul düzeyde olduğu görülmüştür.

Yöntemlere ait uzman kararlarının G kuramı ile incelenmesi sonucunda, üç yöntemin maddelere ait farklılıkları fazla ortaya çıkaramadığı; Ebel yönteminin uzmanlar arası tutarlılık bakımından en düşük sonucu verdiği, onu takiben Angoff yönteminin geldiği ve en yüksek uzman tutarlılığını Yes/No yönteminin verdiği sonucuna ulaşılmıştır. Ayrıca üç yöntemde de hem uzmanlar tarafından maddelere ait güçlük düzeylerinin farklı algılandığı hem de hata varyansının büyük olduğu görülmüştür. Üç yöntemin birlikte ele alındığı G çalışması sonucunda maddeler arası farklılığın ortaya çıkarılmadığı; uzmanlar arasında bir uyumun olduğu ve yöntemler arasında da büyük bir farklılığın olduğu sonucuna ulaşılmıştır. Ayrıca maddelerin bir yöntemden diğer yonteme çok farklı puanlanmadığı; belirli bir uzmanın tüm maddeler için aynı puanlamayı yapmadığı ya da belirli bir maddenin puanlanmasıyla diğer maddelerin puanlaması bakımından bağımlılık olduğu; uzmanların bir yöntemden diğer yonteme kararlı puanlama yapmadıkları, bir başka deyişle uzmanların yöntemlere göre farklı puanlama yaptıkları ve ölçülemeyen hata kaynaklarının büyük olduğu sonucuna ulaşılmıştır.

D çalışması sonuçlarına göre Angoff ve Yes/No yöntemleri ile kesme puanı belirlenirken 17 civarında uzmandan; Ebel yönteminde ise 22 civarında uzmandan karar almanın güvenilirlik için yeterli olacağı; daha fazla sayıda uzmandan karar alındığında yöntemlere ait G ve Phi katsayılarının arttığı, ancak bu katsayılarda yine de büyük bir değişme olmadığı görülmüştür.

Anahtar Sözcükler

Standart Belirleme Yöntemleri, Angoff Yöntemi, Yes/No Yöntemi, Ebel Yöntemi, Kesme Puanı, Genellenebilirlik Kuramı, Puanlayıcı Sayısı

ABSTRACT

GÜNDEĞER, Ceylan. *A Comparison of Angoff, Yes/No and Ebel Standard Setting Methods*, Master's Thesis, Ankara, 2012.

In this study, decisions and cut-off scores obtained from Angoff, Yes/No and Ebel standard setting methods have been compared.

The student data of the research consist of 489 student scores obtained from a achievement test of 20 questions, in the "Division and Fractions" sub-area of 4th grade Mathematics subject, prepared by the researcher. Other data used in the study are 17 experts' decisions, made according to three different standard setting methods for this achievement test.

In the analysis of the data, differences between the percentages of students, considered as successful according to three different standard setting methods, have been tested with the test of the difference between two correlated proportions. The scores of the students have been tailored to binary as successful-unsuccessful according to the cut-off scores obtained from the three methods; and the compliance between them has been examined with Cohen's Kappa statistics. The compliance between the cut-off scores of three methods has been evaluated with Pearson Product-Moment Correlation Coefficient; and the difference between the averages of the cut-off scores has been evaluated with paired samples t test. In order to set the compliance between experts in the methods, Kendall's coefficient of concordance (W) for Angoff and Ebel methods, and Cohen's Kappa statistics for Yes/No method have been calculated. Additionally, variance components and variance percentages of the decisions of methods have been examined with generalizability theory; and the most appropriate number of raters for each method has been set with D study.

At the end of this research, it is seen that there is significant difference on 0,01 error level among percentages of students considered as successful according to these three different standard setting methods. According to the cut-off scores obtained from these three methods, when the scores of the students are tailored to the artificial binary as successful-unsuccessful, it is concluded that there is a high compliance on 0.01 error level between Angoff and Yes/No methods, and Angoff and Ebel methods; and that there is a reasonable compliance on 0.01 error level between Yes/No and Ebel methods. In terms of the cut-off scores of the three methods, it can be said that there is a positive

and moderate relationship between minimum passing scores (MPS) determined Angoff and Yes/No methods; that there is no significant difference between the averages of MPSs of two methods; that there is a positive and high level of relationship between MPSs determined with Angoff and Ebel methods; and that there is significant difference between MPSs of two methods; that there is no significant relationship between MPSs determined with Yes/No and Ebel methods; and that there is significant difference between MPSs of two methods. The compliance of experts in the methods is seen to be at a reasonable level.

As a result of the examination of experts' decisions of the methods with G theory, it is concluded that the three methods could not have demonstrated the differences of the items; that Ebel method gives the lowest result in terms of consistency among experts; that Angoff method comes after; and that Yes/No method gives the highest level of consistency. Additionally, in all of the three methods, it is seen that both the levels of difficulty are differently perceived by experts and the variance of error is great. As a result of G theory, in which three methods are discussed, it is concluded that differences between items could not have been demonstrated; that there is compliance among the experts; and that there is significant difference between the methods. Additionally, it is concluded that the items are not rated differently in a method regarding the other; that a specific expert does not score all of the items in the same way or that there is dependence between scoring of an item and that of other items; that experts do not perform stable scoring from a method to another; in other words, they score differently regarding the methods; and that error sources which cannot be measured are large.

According to D study results, while cut-off score is determined with Angoff and Yes/No methods, it is seen that considering decisions of approximately 17 experts is sufficient for reliability in Angoff and Yes/No methods and 22 experts is sufficient for reliability in Ebel method; and when the decisions are obtained from more experts, methods' G and Phi coefficient increase; but no significant change has been observed in these coefficients.

Key Words

Standard Setting Methods, Angoff Method, Yes/No Method, Ebel Method, Cut-off Score, Generalizability Theory, Rater Number

İÇİNDEKİLER

KABUL VE ONAY.....	i
BİLDİRİM.....	ii
ADAMA.....	iii
TEŞEKKÜR.....	iv
ÖZET.....	v
ABSTRACT.....	vii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ.....	xiii
ŞEKİLLER DİZİNİ.....	xv
BÖLÜM I.....	1
GİRİŞ.....	1
1.1. PROBLEM DURUMU.....	1
1.1.1. Test Merkezli Yöntemlerde Standart Belirleme Süreci.....	3
1.1.1.1. Uzmanların Seçimi.....	3
1.1.1.2. Sınır Grubun Bilgi – Becerilerinin Tanımlanması.....	3
1.1.1.3. Seçilen Yöntemin Kullanımının Uzmanlara Anlatılması.....	4
1.1.1.4. Kararların Toplanması.....	4
1.1.1.5. Geçme Puanının Hesaplanması.....	4
1.1.2. Test Merkezli Standart Belirleme Yöntemleri.....	4
1.1.2.1. Nedelsky Yöntemi.....	5
1.1.2.2. Angoff Yöntemi.....	6

	x
1.1.2.3. Ebel Yöntemi.....	7
1.1.2.4. Yes/No Yöntemi.....	8
1.1.3. Genellenebilirlik Kuramı.....	10
1.1.3.1. Tek Değişkenlik Kaynaklı Evrenler.....	12
1.1.3.2. İki Değişkenlik Kaynaklı Evrenler.....	14
1.1.3.3. Üç veya Daha Çok Değişkenlik Kaynaklı Evrenler.....	15
1.1.3.4. Çaprazlanmış veya Yuvalanmış Değişkenlik Kaynaklarıyla Evrenler.....	15
1.1.3.5. Tesadüfi ve Sabit Değişkenlik Kaynakları.....	17
1.1.3.6. Genellenebilirlik (G) ve Karar (D) Çalışmaları.....	17
1.1.3.7. Göreceli ve Mutlak Kararlar.....	19
1.1.3.8. Genellenebilirlik Katsayısı ve Güvenirlik İndeksi.....	20
1.2. PROBLEM CÜMLESİ.....	23
1.3. ALT PROBLEMLER.....	23
1.4. TANIMLAR VE KISALTMALAR.....	23
1.5. SINIRLILIKLAR.....	24
1.6. SAYILTILAR.....	24
1.7. ARAŞTIRMANIN AMACI VE ÖNEMİ.....	24
1.6. İLGİLİ ARAŞTIRMALAR.....	25
BÖLÜM II.....	33
YÖNTEM.....	33
2.1. ARAŞTIRMANIN TÜRÜ.....	33
2.2. ÇALIŞMA GRUBU.....	33

2.3. VERİLERİN TOPLANMASI VE VERİ TOPLAMA ARAÇLARI.....	34
2.3.1. Öğrencilerden Verilerin Toplanması.....	36
2.3.2. Uzmanlardan Verilen Toplanması.....	37
2.4. VERİLERİN ANALİZİ.....	38
BÖLÜM III.....	43
BULGULAR VE YORUMLAR.....	43
3.1. Alt Problem 1: Angoff, Yes/No ve Ebel standart belirleme yöntemleri ile belirlenen kesme puanlarına göre geçen öğrenci yüzdeleri arasında anlamlı bir farklılık var mıdır?.....	43
3.2. Alt Problem 2: Yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflanma durumları bakımından uyum var mıdır?.....	46
3.3. Alt Problem 3: Yöntemler arasında uzmanlardan elde edilen minimum geçme puanları (MGP) bakımından uyum var mıdır?.....	47
3.4. Alt Problem 4: Yöntemlere ait uzman kararları arasındaki uyum nasıldır?.....	50
3.5. Yöntemlerin varyans bileşenleri nasıldır?.....	50
3.5.1. Angoff yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?.....	51
3.5.2. Yes/No yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?.....	52
3.5.3. Ebel yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?.....	53
3.5.4. Angoff, Yes/No ve Ebel yöntemleri ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?.....	54

3.6. Alt Problem 6: D çalışmasına göre yöntemlere ait en uygun puanlayıcı sayısı nedir?.....	57
3.6.1. Angoff yönteminde en uygun puanlayıcı sayısı nedir?.....	57
3.6.2. Yes/No yönteminde en uygun puanlayıcı sayısı nedir?.....	58
3.6.3. Ebel yönteminde en uygun puanlayıcı sayısı nedir?.....	59
BÖLÜM IV.....	61
SONUÇ VE ÖNERİLER.....	61
4.1. SONUÇLAR.....	61
4.2. ÖNERİLER.....	63
4.2.1. Araştırma Sonuçlarından Çıkan Öneriler.....	63
4.2.2. Araştırmacılara Yönelik Öneriler.....	63
KAYNAKÇA.....	65
EKLER.....	71
EK 1: Başarı Testi.....	72
EK 2: Uzman Görüşleri İçin Araştırma Formu.....	78
EK 3: Angoff Yöntemine Ait Kararlar.....	89
EK 4: Yes/No Yöntemine Ait Kararlar.....	90
EK 5: Madde Bazında Ebel Yöntemine Ait Kararlar.....	91
ÖZGEÇMİŞ.....	92

TABLOLAR DİZİNİ

Tablo 1. Elenmeden Kalan Seçenek Sayısına Göre Doğru Cevap Olasılığı.....	5
Tablo 2. Angoff Yönteminde Kesme Puanının Belirlenmesi.....	6
Tablo 3. Ebel Yönteminin Zorluk ve Uygunluk Düzeyi Aşamaları.....	7
Tablo 4. Yes/No Yönteminde Kesme Puanının Belirlenmesi.....	8
Tablo 5. Tek Değişkenlik Kaynaklı Ölçümler İçin Değişkenlik Kaynakları.....	13
Tablo 6. İki Değişkenlik Kaynaklı Ölçümler İçin Değişkenlik Kaynakları.....	15
Tablo 7. Araştırmaya Katılan Okullar ve Öğrenci Sayıları.....	34
Tablo 8. Ön Uygulama Sonuçlarına Göre Madde İstatistikleri.....	35
Tablo 9. Nihai Testin Kestirilen Özellikleri.....	36
Tablo 10. Nihai Uygulama Sonuçlarına Göre Test İstatistikleri.....	37
Tablo 11. Frekans ve Oranlara Göre İstatistiksel Önem Testi.....	39
Tablo 12. Tek Değişkenli Çaprazlanmış Modelle Varyans Bileşenlerinin Kestirilmesi.....	41
Tablo 13. İki Değişkenli Çaprazlanmış Modelle Varyans Bileşenlerinin Kestirilmesi.....	42
Tablo 14. Yöntemler İçin Elde Edilen MGP'ler.....	43
Tablo 15. Yöntemlere Ait Kesme Puanları, Başarılı Kabul Edilen Öğrenci Sayıları ve Yüzdeleri.....	44
Tablo 16. Kesme Puanları, Başarılı Kabul Edilen Öğrenci Sayıları - Yüzdeleri ve z Değerleri.....	45
Tablo 17. Yöntemler Arasındaki Uyum İçin Hesaplanan Korelasyon Katsayıları.....	46
Tablo 18. MGP'lere Ait Betimsel İstatistikler ve Normallik Testi Sonuçları.....	47
Tablo 19. MGP'ler Arasındaki Pearson Korelasyon Katsayıları.....	48

Tablo 20. MGP Ortalamaları Arasındaki Farkın Testi.....	49
Tablo 21. Angoff Yöntemine Ait Kestirilen Varyans ve Toplam Varyansı Açıklama Yüzdeleri.....	51
Tablo 22. Yes/No Yöntemine Ait Kestirilen Varyans ve Toplam Varyansı Açıklama Yüzdeleri.....	52
Tablo 23. Ebel Yöntemine Ait Kestirilen Varyans ve Toplam Varyansı Açıklama Yüzdeleri.....	53
Tablo 24. Üç Yönteme Ait Kestirilen Varyans ve Toplam Varyansı Açıklama Yüzdeleri.....	55
Tablo 25. Angoff Yönteminde Farklı Uzman Sayılarına Ait G ve Phi Katsayıları.....	57
Tablo 26. Yes/No Yönteminde Farklı Uzman Sayılarına Ait G ve Phi Katsayıları.....	58
Tablo 27. Ebel Yönteminde Farklı Uzman Sayılarına Ait G ve Phi Katsayıları.....	59

ŞEKİLLER DİZİNİ

Şekil 1. G Kuramının Kökenleri ve Kavramsal Yapısı.....	11
Şekil 2. Tek Değişkenlik Kaynaklı Ölçümler İçin Çaprazlanmış $b \times m$ Desenine Ait Değişkenlik Kaynaklarının ve Varyans Bileşenlerinin Venn Diyagramlarıyla Gösterimi.....	14

BÖLÜM I

GİRİŞ

Bu bölümde araştırmanın problem durumuna, amacına, önemine, problem cümlesine, alt problemlerine, sayıltı ve sınırlılıklarına yer verilmiştir.

1.1. PROBLEM DURUMU

Tüm dünyada olduğu gibi ülkemizde de öğrenci başarısının belirlenmesinde, işe alımlarda veya öğrenci seçimlerinde çeşitli testlerden yararlanılmaktadır. Bu testler, amaca uygun olarak, bireyin istenilen özelliklere ne kadar sahip olduğunu ortaya çıkaracak şekilde hazırlanmaktadır. Bireyin istenilen özelliklere ne derece sahip olduğu testlerle belirlendiği gibi, bireyler hakkındaki kararlar da yine test sonuçları dikkate alınarak verilmektedir.

Ülkemizde eğitim alanında her yıl Lisans Yerleştirme Sınavı, Kamu Personeli Seçme Sınavı, Kamu Personeli Dil Sınavı, Akademik Personel ve Lisansüstü Eğitim Giriş Sınavı vb. pek çok sınav yapılmaktadır. Bu sınavlardan bazıları (örneğin LYS) sınava girenleri sıralamakta ve bu sıralamaya göre öğrencileri yerleştirmekte iken; bazıları ise (örneğin KPDS) belirli bir puanın üstünde puan alanları “başarılı” kabul etmektedir. Kısaca LYS gibi sınavlarda bağıl değerlendirme kullanılırken; KPDS gibi sınavlarda ise mutlak değerlendirme kullanılmaktadır.

Bağıl değerlendirme, önceden belirlenmemiş, ölçme işlemi sonrasında ölçme sonuçlarına dayalı olarak elde edilen bağıl ölçüte dayalı olarak yapılan değerlendirme türüdür. Mutlak değerlendirme ise değerlendirmede kullanılacak ölçütün ölçme sonuçlarına bağlı olmadan elde edilmesi ve değerlendirmede kullanılması durumudur. Kısacası bağıl değerlendirmede değerlendirilecek grubun puanlarına bağlı olarak bir ölçüt belirlenirken; mutlak değerlendirmede ölçütün grubun puanlarından bağımsız olarak belirlenmesi söz konusudur.

Değerlendirmelerde kullanılan ölçütler, standart belirleme yöntemleri ile belirlenebilmektedir. Standart, “*Ne kadar yeterlidir ?*” sorusunun cevabıdır. Kesme puanı (ölçüt) belirlenirken aslında test performansı için bir standart belirlenmiş olur. Standartlar bağıl (*relatively*) ya da mutlak (*absolute*) olabilir. Bağıl standartlar bireyler arası karşılaştırmalara bağılı iken; mutlak standartlar grup içi karşılaştırmalara bağılı olmaktan ziyade belirlenen bir ölçütle ilgilidir (Livingston ve Zieky, 1982, s. 10 – 11). Bu iki kavramı aşağıdaki iki örnekle açıklamak mümkündür:

“*Soruların en az %55’ini doğru cevaplayanlar geçecek.*” ifadesi mutlak bir standart oluştururken; “*Puanı grup ortalamasının üstünde olanlar geçecek.*” ifadesi bağıl bir standart oluşturmaktadır. Aynı şekilde, “*Soruların %60’ını cevaplayanlar geçecek.*” ifadesi mutlak bir standart oluştururken; “*Grubun %60’ını geride bırakanlar geçecek.*” ifadesi gruba bağımlılık belirtmesiyle bağıl bir standart oluşturmaktadır.

Örneklerden yola çıkarak, testi alan her bir kişinin puanı, testi alan diğer kişilerin puanlarıyla karşılaştırıldığı için bağıl standartların gruba bağılı olduğunu ve norm temelli olarak da adlandırıldığını söyleyebiliriz. Bu standart türünde testi alan kişilerin puanı ne kadar artarsa, standart da o kadar yükselecektir. Bağıl standartların aksine, mutlak standartlar gruba (testi alanların performansına) bağılı değildir ve ölçüt temelli standartlar olarak da bilinirler. Ölçüt temelli standartlarda, testi alan her bir birey için testi alan diğer kişilerin puanları önemli değildir, çünkü onların puanları standardı etkilememiştir (Livingston ve Zieky, 1982, s. 11).

Bu açılarından standart, bir “*ölçüt*” ifade eder. Ölçüt, geçme puanının minimum sınırı olarak düşünülebilir. Geçme puanı ise istenilen performans düzeyinde olan öğrenci ile yeterli düzeyde olmayan öğrenciyi ayırt edebilen uygun performans noktasını gösterir. İşte bu geçme puanını uygun şekilde belirleme, bir başka deyişle öğrencilerin geçti-kaldı ve/veya başarılı-başarısız durumlarına karar vermek için bir ölçüt belirleme işi standart belirleme sürecini gerektirir. Standart belirleme süreci, kişileri performans düzeylerine göre ayıran performans standartlarını, test puanları cetvelinde rakamsal yani işevuruk hâle getirme süreci olarak düşünülebilir (Hambleton, 2001, s. 92). Bu süreç yöntemden yönetime farklılık göstermektedir.

Son yıllarda alanyazında otuzdan fazla standart belirleme yöntemi betimlenmesine karşın, bu yöntemleri genel olarak üç ana başlık altında toplamak mümkündür (Crocker ve Algina, 1986, s. 411):

- i. Bütünsel Etkiye Dayanan Yöntemler
- ii. Test Maddelerine Dayanan (*Test Merkezli*) Yöntemler
- iii. Testi Alanların Performansına Dayanan (*Performans Merkezli*) Yöntemler

Bu araştırma kapsamında test merkezli yöntemlerden Angoff, Yes/No ve Ebel yöntemleri kullanılmıştır. Bu nedenle öncelikle test merkezli yöntemlerde standart belirleme sürecinin genel olarak nasıl işlediğine dair bilgi vermenin uygun olacağı düşünülmüştür.

1.1.1. Test Merkezli Yöntemlerde Standart Belirleme Süreci

Test merkezli yöntemler, ülkemizde sayıca az olsalar da, özellikle yurtdışında birçok araştırma durumuna konu olmuşlardır. Test merkezli yöntemler ile standart belirlenirken aşağıda bahsedilen sürece dikkat edilmesi önerilmektedir (Livingston ve Zieky, 1982, s. 15 – 17):

1.1.1.1. Uzmanların Seçimi

Uzmanlar, “*Ne kadar yeterlidir ?*” sorusuna cevap verecek olan, diğer bir deyişle standardı belirleyecek olan uzman grubudur. Bu uzman grubundaki kişiler, ele alınacak testin alanıyla ilgili yeterli donanıma ve problem çözme becerisine sahip, öğretmenler gibi eğitimle ilgili kişiler olmalıdırlar. Ayrıca standart belirleme sürecinde de testin amacını göz önünde bulundurulmalıdır.

1.1.1.2. Sınır Grubun Bilgi – Becerilerinin Tanımlanması

Seçilen uzmanların ilk görevi, sınır grubun bilgi – becerilerini kendi ifadeleriyle tanımlamaktır. Bu süreç içinde iletişim çok önemlidir. Öncelikle testin neyi ölçtüğü,

neyi amaçladığı, testten elde edilen puanların nasıl kullanılacağı uzmanlara detaylarıyla anlatılmalıdır (Örneğin test sınıf içi başarı testi ise standart belirleme sürecinde bu durum dikkate alınmalıdır). Ardından uzmanlardan sınır grubu tanımlamaları istenir. Ne kadar sağlıklı tanımlamalar yapılırsa geçme puanı da o kadar güvenilir olacaktır.

1.1.1.3. Seçilen Yöntemin Kullanımının Uzmanlara Anlatılması

Bu aşamada uzmanlara, seçilen yöntemin uygulanışı ve geçme puanının belirlenişi anlatılmalı, yöntem hakkında detaylı bilgi verilmeli, uzmanların yöntemi anladıklarından emin olunmalıdır. Hatta gerekirse bir ön uygulama ile uzmanlar bu sürece yatkın hâle getirilmelidir.

1.1.1.4. Kararların Toplanması

Her uzman bireysel olarak, belirlenen yöneme uygun şekilde testle ilgili kararlarını verir. Bu aşama her yöntem için farklılık göstermektedir.

1.1.1.5. Geçme Puanının Hesaplanması

Her uzmana göre minimum geçme puanı (MGP) hesaplanır. Bu puanların ortalaması alınarak testin nihai geçme puanı belirlenir. Ortalamanın yanında, “ortanca” veya “budanmış ortalama” (trimmed mean) denilen ve aşırı uç puanları hesaplama dışı bırakan özel bir ortalama geçme puanı belirlemede kullanılabilir.

1.1.2. Test Merkezli Standart Belirleme Yöntemleri

Araştırmada kullanılan üç test merkezli yöntem olan Angoff, Ebel ve Yes/No yöntemleri, ayrıca yine test merkezli yöntemlerden biri olan ve oldukça yaygın kullanıma sahip olan Nedelsky yöntemi aşağıda detaylı bir şekilde açıklanmaktadır.

1.1.2.1. Nedelsky Yöntemi

1954 yılında Leo Nedelsky tarafından önerilmiş olan standart belirleme yöntemidir. Bu yöntem *yalnızca* çoktan seçmeli testlerde kullanılabilir. Çünkü yöntem, “*Öğrenci testi çözerken seçenekleri eleme sürecini izler, yanlış olduğunu düşündüğü seçenekleri eleyerek doğru cevaba ulaşır.*” fikrini temele almaktadır (Livingston ve Zieky, 1982, s. 17).

Uzmanlar bu yöntemde bireysel olarak, minimum yeterlilik düzeyindeki öğrencinin çoktan seçmeli test maddesindeki seçeneklerden eleyebileceği yanlış seçenek sayısını tahmin etmeye çalışırlar. Uzmanların, sınır grup öğrencisinin eleyebileceğini düşündüğü seçenek sayısına göre kalan seçeneklerden yola çıkılarak yüzdeler elde edilir. Her bir madde için elde edilen bu yüzdeler toplanarak her bir uzmana ait MGP hesaplanır. Uzmanlara ait MGP’lerin ortalaması (ortancası veya trimmed mean) ile de testin nihai geçme puanı belirlenir (Livingston ve Zieky, 1982, s. 17 – 24). Nedelsky yönteminde 4 seçenekli bir soru için elenmeden kalan seçenek sayısına göre doğru cevap olasılıkları Tablo 1’de gösterilmiştir.

Tablo 1: Elenmeden Kalan Seçenek Sayısına Göre Doğru Cevap Olasılığı

Elenmeden Kalan Seçenek Sayısı	Maddenin Doğru Cevaplanma Olasılığı
4 seçenek	0,25
3 seçenek	0,33
2 seçenek	0,50
1 seçenek	1,00

Tablo 1’den de anlaşıldığı üzere uzmanın verdiği karara göre maddelerin cevaplanma olasılığı değişecektir. Örneğin uzman 4 seçenekli bir madde için, sınır grup öğrencileri tarafından hiçbir seçeneğin elenmeyeceğini düşünürse o maddeye ait doğru cevaplanma olasılığı 0,25; 1 seçeneğin eleneceğini düşünürse geriye kalan seçenek sayısı 3 olduğundan maddenin doğru cevaplanma olasılığı 0,33 olacaktır.

1.1.2.2. Angoff Yöntemi

1971 yılında William H. Angoff tarafından önerilmiş olan standart belirleme yöntemidir (Zieky, 2001, s. 22). Angoff yöntemi benzer diğer yöntemlere göre sürecin uzmanlara açıklanması, verilerin toplanması ve analizi bakımından daha kolay olan, daha sık kullanılan bir yöntemdir (Mills ve Melican, 1988).

Angoff yönteminde madde bir bütün olarak düşünülür ve sınır grubun maddeye doğru cevap verme olasılığına odaklanılır. Bu nedenle yöntem tüm sınav türlerinde kullanılabilir. Sınır grubun maddelere doğru cevap verme olasılığı 0 (sıfır) ve 1 arasında değerler alır. Sorular kolaylaştıkça bu olasılık artacaktır. Eğer herhangi bir çoktan seçmeli test maddesi için olasılık sıfır olarak düşünüldüyse, o test maddesine ait şans başarısı olan “*1/seçenek*” sayısı olasılık olarak kabul edilmelidir. Ayrıca uzmanlar yöntemde zorlanırlarsa 100 kişilik bir sınır grup hayal ederek bunlardan kaçının maddeye doğru cevap vereceğini belirleyebilirler (Livingston ve Zieky, 1982, s. 24 – 26). Angoff yönteminde kesme puanı belirlenmesine yönelik bir örnek aşağıda Tablo 2’de verilmiştir.

Tablo 2: Angoff Yönteminde Kesme Puanının Belirlenmesi

Uzmanlar	Maddeler					MGP
	1	2	3	4	5	
1	%60	%70	%90	%50	%50	64
2	%55	%100	%70	%50	%40	63
3	%70	%90	%55	%70	%20	61
4	%60	%80	%80	%70	%30	64
Kesme Puanı						63

Tablo 2’de de görüldüğü gibi uzmanların maddelere verdikleri sınır grubun maddeyi doğru cevaplama yüzdeleri toplanarak her bir uzmana ait geçme puanı (MGP) hesaplanır. Uzmanların belirledikleri geçme puanlarının (MGP’lerin) ortalaması ile (ortanca veya trimmed mean de olabilir) testin nihai geçme puanı (kesme puanı) saptanır.

Colton ve Hecht (1981) tarafından Angoff yönteminin diğer yöntemlerden daha üstün yanlarının olduğu ileri sürülmüştür (Aktaran: Cizek ve Bunch, 2007, s. 82). Yöntemin yurtdışında genellikle lisans ve sertifika programlarının yanında sağlık bilimleri alanındaki çoktan seçmeli maddelerden oluşan testlerde kullanıldığı görülmektedir.

1.1.2.3. Ebel Yöntemi

1972 yılında R. L. Ebel tarafından önerilmiş standart belirleme yöntemidir. Yöntem genellikle sağlık bilimleri alanında ve en fazla da sınıf içi başarı testlerinde standart belirleme amacıyla kullanılmaktadır (Cizek ve Bunch, 2007, s. 75).

Test maddelerine dayanan diğer yöntemler gibi Ebel yöntemi de uzmanların her bir test maddesi hakkında karar vermelerini gerektirir. Ancak yöntemde diğer yöntemlerden farklı olarak uzmanların her bir madde için iki farklı karar vermeleri gerekmektedir. Bu kararlardan biri her madde için zorluk ve uygunluk düzeyinin tahmin edilmesidir (Cizek ve Bunch, 2007, s. 75 – 76). Bir başka deyişle yöntemde maddelerin *Zorluk Düzeyi* ve *Uygunluk Düzeyi* olmak üzere iki aşamada sınıflandırılması söz konusudur. Bu iki aşama kendi içinde “*Zor, Orta, Kolay*” ve “*Gerekli, Önemli, Kabul Edilebilir, Tartışılabilir*” olarak derecelendirilmiştir (Ebel, 1972, s. 493). Sonuç olarak ortaya aşağıdaki Tablo 3 gibi 3 x 4’lük bir tablo çıkmaktadır.

Tablo 3: Ebel Yönteminin Zorluk ve Uygunluk Düzeyi Aşamaları

Uygunluk	Zorluk		
	Kolay	Orta	Zor
Gerekli	Maddeler: Yüzde:	Maddeler: Yüzde:	Maddeler: Yüzde:
Önemli	Maddeler: Yüzde:	Maddeler: Yüzde:	Maddeler: Yüzde:
Kabul Edilebilir	Maddeler: Yüzde:	Maddeler: Yüzde:	Maddeler: Yüzde:
Tartışılabilir	Maddeler: Yüzde:	Maddeler: Yüzde:	Maddeler: Yüzde:

Ebel yönteminde ilk olarak, uzmanlardan maddeleri *sadece* testin yapısını dikkate alarak yukarıdaki sınıflama tablosuna uygun şekilde yerleştirmeleri istenir. Sonrasında ise sınır grup öğrencilerini dikkate alarak bu sınıflamalara ait doğru cevaplanma yüzdelerinin belirlenmesi gerekmektedir. Bu iki aşama tamamlandıktan sonra, her hücrede bulunan madde sayısı ile o hücreye ait yüzde çarpılarak her hücrenin puanı belirlenir ve belirlenen bu puanlar toplanarak uzmanlara ait geçme puanları (MGP) hesaplanır. Her bir uzmana ait geçme puanlarının ortalaması (ortanca veya trimmed mean de olabilir) testin nihai geçme puanını verir (Livingston ve Zieky, 1982, s. 26 – 29).

1.1.2.4. Yes/No Yöntemi

1997 yılında J. C. Impara ve B. S. Plake'in geliştirmiş olduğu ve Angoff yönteminin de bir varyasyonu olan standart belirleme yöntemidir. Yöntem aynı zamanda “*Doğrudan Sınır Yöntemi*” olarak da bilinmektedir (Cizek ve Bunch, 2007, s. 88 – 89).

Yes/No yönteminde uzmanlardan testteki maddeleri inceleyerek, sınır gruptaki bir öğrencinin maddeyi doğru cevaplayıp cevaplamayacağına karar vermeleri istenir. Uzman tarafından, sınır grup öğrencisinin maddeyi doğru cevaplayacağı düşünülüyorsa maddeye 1, yanlış cevaplayacağı düşünülüyorsa maddeye 0 (sıfır) puan verilerek, her bir uzmana ait geçme puanı (MGP) hesaplanır. MGP'lerin ortalaması testin nihai geçme puanı kabul edilir (Zieky, 2001, s. 22). Yes/No yönteminde kesme puanı belirlenmesine yönelik bir örnek Tablo 4'te verilmiştir.

Tablo 4: Yes/No Yönteminde Kesme Puanının Belirlenmesi

Uzmanlar	Maddeler					MGP
	1	2	3	4	5	
1	1	0	0	1	0	2 (40)
2	1	1	0	1	1	4 (80)
3	1	1	1	0	1	4 (80)
4	1	1	0	0	0	2 (40)
Kesme Puanı						3 (60)

Yes/No yöntemi geleneksel Angoff yöntemine göre daha kolaydır. Angoff yönteminde uzmanların sınır grup öğrencilerinden yüzde kaçının maddeye doğru cevap vereceğini tahmin etmeleri istenir ve bu görev zordur (Cizek ve Bunch, 2007, s. 89 – 90).

Impara ve Plake (1998, s. 69 – 81) Yes/No yönteminin Angoff yöntemindeki yüzde tahmini görevinde karşılaşılan zorlukları iyileştirdiğini ortaya koymuşlardır. Araştırmaya göre, uzmanlar Yes/No yönteminin karar verme sürecini geleneksel Angoff yöntemindeki tahmin sürecine göre daha açık ve daha kolay bulmuşlardır ancak sonuçlar Yes/No yönteminin daha az kararlı olduğunu göstermektedir.

Alanyazında bütün testler için genel geçer bir “*en iyi yöntem*” tanımlanmamıştır. Test merkezli yöntemlerin hepsi uzmanların karar verme sürecini gerektirir ve bu kararlar öznelidir (Downing ve diğerleri, 2006, s. 57). Ancak kararların elde edilmesinde tek bir birey yerine uzman grubu yer aldığı için güvenilirlik sağlanabilir. Ayrıca bu yöntemler testin uygulanmasından önce de sonra da kullanılabilir. Yöntemlerin avantajları kadar dezavantajları da vardır. Kararların varsayımlara dayanması ve kararlardaki aşırı farklılıklar birer sınırlılık olarak düşünülebilir. Uzman kararlarında tam anlamıyla uzlaşma sağlamak zordur. Belirlenen standartlar, uygulamadaki grup başarısından çok yüksek veya çok düşük çıkabilir. Bu nedenle belirlenen standarda göre başarılı olması/sayılması gereken öğrenci başarısız; başarısız olması/sayılması gereken öğrenci de başarılı olabilir/sayılabilir ki bu da hata anlamına gelmektedir.

Standart belirleme yöntemleri ile elde edilen ölçütlere göre öğrenci başarıları, öğrencilerin geçti-kaldı ve/veya başarılı-başarısız olma durumları belirlenmektedir. Öğrenci hakkında karar verme aşamasında bir ölçüt olan standartlar, aynı test için farklı standart belirleme yöntemlerine göre farklılık gösterebilmektedir. Farklı standart belirleme yöntemleri arasında nasıl bir ilişki vardır? Belirlenen bu standartların güvenilirliği nedir? Standartları belirleyen uzman grubunun birbiriyle uyumu nasıldır? Araştırma kapsamında Genellenebilirlik (G) kuramı ile bu sorulara cevap aranacaktır. Bu nedenle G kuramından bahsetmek yerinde olacaktır.

1.1.3. Genellenabilirlik (G) Kuramı

G kuramı, performans ölçümlerinin güvenilirliği ile ilgili istatistiksel bir kuramdır (Shavelson ve Webb, 1991, s. 1). Kuram birçok ölçme konusunu ele almada, geniş bir kavramsal çerçeve ve güçlü bir istatistiksel süreç sunmuştur (Brennan, 2001, s. 2).

Kuram varyans analizi (ANOVA) süreçlerinin ölçme sonuçlarına uygulanması nedeniyle Klasik test kuramının (KTK) bir uzantısı olarak düşünülebilir. KTK, gözlenen ölçümlerin, gerçek puan (T) ve tesadüfi hatadan (E) oluştuğunu varsayar (Brennan, 2001, s. 2). KTK'da gözlenen puan içinde, gerçek puan ile ölçmede yapılan tesadüfi hatadan gelen bir miktar hata puanı bulunur. Bu açıklama KTK'nın temel sayılıtısını oluşturur. Gözlenen puan, gerçek puan ve hata puanı cinsinden

$$X = T + E \quad (1)$$

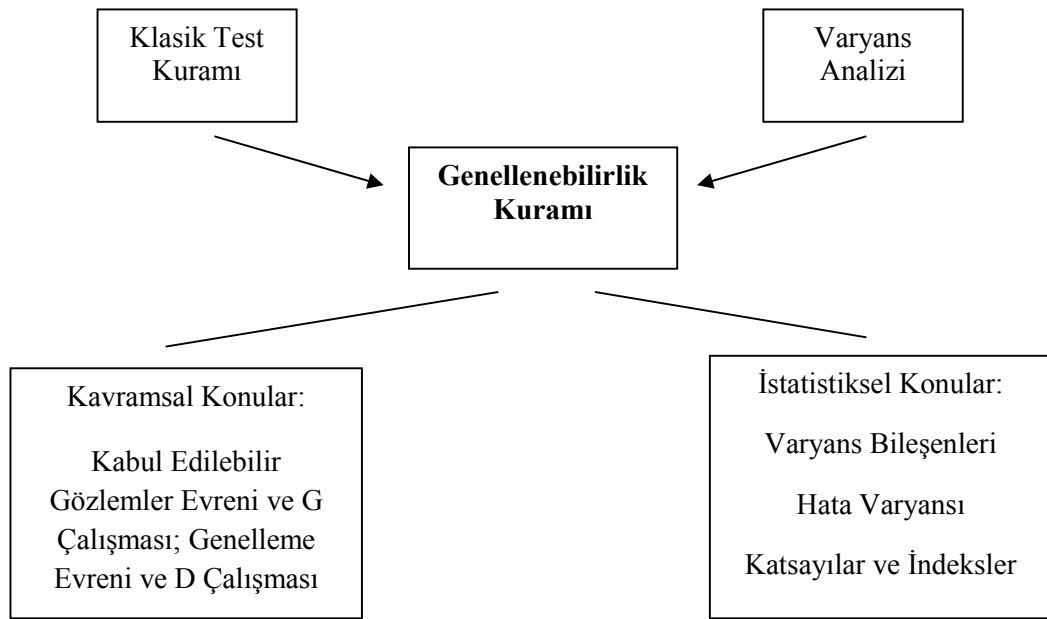
ifadesiyle yazılır. Bu ifadeye KTK'nın temel denklemi denir (Baykul, 2000, s. 113).

Tesadüfi hata (E) ölçme sonuçlarına gelişi güzel karışan, kaynağı, miktarı ve yönü kesin olarak bilinmeyen hatalardır (Atılğan, 2009, s.26; Baykul, 2000, s. 107; Tekin, 1991, s. 56; Turgut, 1993, s.29). Bu nedenle KTK'nın herhangi bir uygulaması çoklu hata kaynaklarını net bir şekilde ayırt edemez (Brennan, 2001, s. 2).

G Kuramı, araştırmacıya KTK'da ayırt edilmemiş olan (E) çoklu hata kaynaklarını ANOVA yöntemini kullanarak belirleme izni verir. KTK ve ANOVA, G kuramının ataları gibi görünse de G kuramı kökenlerinin sahip olduğundan daha fazla bilgi içerir. G kuramının en önemli yanı kavramsal yapısıdır (Brennan, 2001, s. 3 – 4).

Klasik test kuramında kararın dayandığı puan, aynı amaca hizmet eden birçok puandan yalnızca biridir. Tek puan, tek bir test oturumunda bir kerede elde edilir. Bireyin ortalama puanı, kabul edilebilir tüm zamanlar, test formları ve oturumlar ile eşleşmeye elverişli değildir. Bu nedenle de güvenilir değildir. Bireyin puanı genellikle diğer şartlarda, diğer test formlarında veya diğer oturumlarda farklı olacaktır. Bunlardan

hangisi en önemli hata kaynağıdır? Klasik test kuramı bir seferde yalnızca bir hata kaynağın tahmin edebilmektedir. G kuramının avantajı, birden fazla hata kaynağını tek bir analizde ayrıştırarak tahmin edebilmesidir. G kuramı, güvenilirlik düzeyini etkileyen bileşenler hakkında özet bilgi ve klasik test kuramında güvenilirlik katsayısı olarak bilinen genellenebilirlik katsayısını vermektedir. Ayrıca kuram araştırmacıya D çalışması ile güvenilir puanı gözlemek için ne kadar oturum-test formu ve puanlayıcı gerektiğini belirleme olanağı da sunmaktadır (Shavelson ve Webb, 1991, s. 1 – 2).



Şekil 1: G Kuramının Kökenleri ve Kavramsal Yapısı (Brennan, 2001, s. 5)

Güvenirlik, bireyin bir test ya da başka ölçümlerdeki gözlenen puanından bütün olası koşullar altında alabileceği ortalama puanına kadar genellenebilirliğin doğruluğuna dayanır. Farklı durumlarda ölçülen ve bireysel olarak elde edilen puanlar arasındaki herhangi bir farklılığın bir ya da daha fazla hata kaynağından geldiği varsayılır. G kuramı araştırmacıya farklı durumlar için güvenilirlik katsayısı hesaplama olanağı da sağlar. Örneğin, bir öğrenciye ait puan 95'ten yüksek olsun. Öğretmeni, bu öğrenciye ait performansın mutlak düzeyini sınavından bağımsız olarak bilmek isterse ne olur? G kuramı bu tür durumlar için de güvenilirlik katsayısı hakkında bilgi vermektedir (Shavelson ve Webb, 1991, s. 2).

Kısacası G kuramı dört farklı açıdan KTK'nın genişletilmiş bir hâlidir (Shavelson ve Webb, 1991, s. 2):

- i. G kuramı çoklu varyans kaynaklarını tek bir analizde ele alır.
- ii. Her bir varyans kaynağının büyüklüğünün hesaplanmasını sağlar.
- iii. Hem bireylerin performanslarına dayalı göreceli kararların hem de bireylerin performanslarıyla ilgili mutlak kararların alınmasına ilişkin iki farklı güvenilirlik katsayısının (G ve Phi katsayıları) hesaplanmasına olanak verir.
- iv. Belirli bir amaca bağlı olarak ölçme hatasının en aza indirgenebileceği ölçme durumlarının (senaryolarının) düzenlenmesine (D çalışmaları) imkân tanır.

G Kuramında bir ölçme durumları seti, değişkenlik kaynağı (facet) olarak adlandırılır. Bir başka deyişle değişkenlik kaynağı, ölçme hatasının olası kaynağı olarak tanımlanabilir. Örneğin, fabrika çalışanlarının performansının iki uzman tarafından ağır-orta-hafif iş yükü olarak puanlandığını varsayalım. Burada değişkenlik kaynakları uzmanlar ve iş gücüdür (Crocker ve Algina, 1986, s. 159). G kuramının daha iyi anlaşılması adına aşağıda bu kavramlara detaylıca yer verilmiştir.

1.1.3.1. Tek Değişkenlik Kaynaklı Evrenler

Tek değişkenlik kaynaklı evrenler, tek bir ölçme hata kaynağı ile tanımlanırlar. G kuramına göre ölçme sonuçları, kabul edilebilir gözlemler evreninin örnekleimidir. Araştırmacı bir teste ait maddeleri evrene genellemeyi isterse değişkenlik kaynağı maddeler, madde evreni de tüm maddelerin bulunduğu evren olacaktır. Araştırmacı bir test formunu evrene genellemeyi isterse aynı şekilde değişkenlik kaynağı test formu, test formu evreni de tüm test formlarından (örneğin son 15 yılda geliştirilen tüm testlerden) oluşan evren olacaktır (Shavelson ve Webb, 1991, s. 3).

Tek değişkenli desenlerin dört değişkenlik kaynağı vardır. Bunlardan biri öğrencilerin başarıları arasındaki sistematik farklılıktan kaynaklanır. Bu değişkenlik kaynağına ölçmenin amacı (*object of the measurement*) denir. Ölçmenin amaçları arasındaki değişkenlik, sosyal bilimlerde genellikle bireylerin bilgi, beceri gibi özelliklerindeki farklılıklarına karşılık gelir (Shavelson ve Webb, 1991, s. 4).

Değişkenliğin ikinci kaynağı test maddelerinin güçlüklerindeki farklılıklardan kaynaklanır. Bazı test maddeleri zordur, bazıları kolaydır ve bazıları da orta güçlüktedir (Shavelson ve Webb, 1991, s. 5).

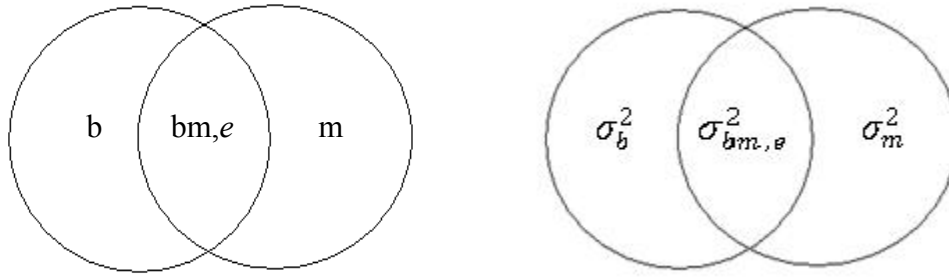
Üçüncü değişkenlik kaynağı öğrencilerin teste getirmiş oldukları eğitimsel ve deneysel geçmişleridir. Örneğin farelerle ilgili bir madde, fare (hamster) beslemiş bir öğrenci için diğer öğrencilere göre daha kolay olacaktır. Öğrenciler arasındaki bu farklılık, varyans analizinde birey-madde etkileşimi olarak geçer. Bireyin geçmişi ve birey ile belirli bir madde arasındaki bu eşleşme, testten elde edilen öğrenci puanının evrene genellenmesini güçleştirir (Shavelson ve Webb, 1991, s. 5 – 6).

Dördüncü değişkenlik kaynağı öğrenci dikkati gibi rastgelelikten, farklı öğrencilerin farklı günlerde aynı testi almaları gibi diğer sistematik ama tanımlanamayan/bilinmeyen değişkenlik kaynaklarından ya da her ikisinden de kaynaklanabilir (Shavelson ve Webb, 1991, s. 6).

Özetle dört değişkenlik kaynağı, (a) ölçme amaçları (bireyler) arasındaki farklılıklar, (b) madde güçlüklerindeki farklılıklar, (c) birey-madde etkileşimi ve (d) rastgele ya da tanımlanamayan olaylar olarak tanımlanabilir. Üçüncü ve dördüncü değişkenlik kaynakları “artık” (*residual*) olarak beraber düşünülür: Birey-madde etkileşimi ($b \times i$) diğer tanımlanamayan kaynaklarla beraber ele alınır ve “ e ” ile gösterilir (Shavelson ve Webb, 1991, s. 6). Tek değişkenlik kaynaklı ölçümler için değişkenlik kaynakları aşağıda Tablo 5’te ve Şekil 2’de gösterilmektedir.

Tablo 5: Tek Değişkenlik Kaynaklı Ölçümler İçin Değişkenlik Kaynakları

Değişkenlik Kaynağı	Değişkenlik Türü	Varyans Gösterimi
Birey (b)	Evren puanı	σ_b^2
Madde (m)	Durumlar	σ_m^2
b x m etkileşimi	Artık	$\sigma_{bm,e}^2$



Şekil 2: Tek Değişkenlik Kaynaklı Ölçümler İçin Çaprazlanmış $b \times m$ Desenine Ait Değişkenlik Kaynaklarının ve Varyans Bileşenlerinin Venn Diyagramlarıyla Gösterimi (Shavelson ve Webb, 1991, s. 21).

1.1.3.2. İki Değişkenlik Kaynaklı Evrenler

Sosyal bilimlerdeki ölçmeler karmaşıktır ve bu nedenle birden fazla değişkenlik kaynağı içerirler (Shavelson ve Webb, 1991, s. 6).

Davranış gözlenmesinde değişkenlik kaynağı, puanlayıcılar ve zaman olacaktır. Puanlayıcılar arasındaki tutarsızlık ne kadar büyük olursa, bir gözlemcinin puanının evrene genellenmesi de o kadar hatalı olur. Ayrıca, davranışların belirli bir zamandan diğer bir zamana farklılık gösterip göstermediğinin anlaşılması bakımından, davranış gözlemlerinin tekrarlanması araştırmalarda sıklıkla izlenen bir yöntemdir. Puanlayıcı, bireyleri birkaç sefer izler. Bireylerin farklı zamanlarda sergiledikleri davranışları arasındaki tutarsızlıklar ne kadar büyük olursa, davranışların bütün zamanlarda sergilenen davranışlar evrenine genellenmesi de o kadar zor olur (Shavelson ve Webb, 1991, s. 7).

İki değişkenlik kaynaklı desenlerin, ölçme değişkenlik kaynaklarına bağlı olarak 6 değişkenlik kaynağı bulunur. Aşağıda Tablo 6’da iki değişkenlik kaynaklı ölçümler için değişkenlik kaynakları yer almaktadır. Değişkenliğin son kaynağı, birey-puanlayıcı-zaman üçlüsünün özgün kombinasyonu, varyansın ölçülmeyen kaynakları ve rastgele olaylar, artıktır. Tek değişkenlik kaynaklı desenlerde kullandığımız gibi “e” ölçülmemiş ve rastgele varyans kaynağıdır. Tek değişkenli hata kaynaklarında olduğu gibi varyans bileşeni, her kaynağın ne kadar önemli olduğunu gösterir (Shavelson ve Webb, 1991, s. 8 – 10).

Tablo 6: İki Değişkenlik Kaynaklı Ölçmeler İçin Değişkenlik Kaynakları

Değişkenlik Kaynağı	Değişkenlik Türü	Varyans Gösterimi
Birey (b)	Evren-puan varyansı (ölçmenin amacı)	σ_b^2
Puanlayıcı (p)	Bütün bireyler üstündeki sabit etki	σ_p^2
Zaman (z)	Bütün bireyler üstündeki sabit etki	σ_z^2
b x p	Bireylerin puanlanması sırasında puanlayıcılar arası tutarsızlıklar	σ_{bp}^2
b x z	Bir zamandan diğerine bireyin davranışlarındaki tutarsızlıklar	σ_{bz}^2
p x z	Sabit etki	σ_{pz}^2
b x p x z, e	Artan/Kalan	$\sigma_{b,pz,e}^2$

1.1.3.3. Üç veya Daha Çok Değişkenlik Kaynaklı Evrenler

Sosyal bilimlerdeki ölçmelerin karmaşıklığı her zaman iki değişkenlik kaynağı ile ele almaya yeterli olmayabilir. Araştırmacı madde, zaman ve test uygulayıcısı gibi üç (ya da daha fazla) değişkenlik kaynağına genellemek isteyebilir. Örneğin test performansı, test uygulayıcısına ait yönergenin açıklığına bağlı olarak değişebilir. Bu durumda üç değişkenlik kaynağı maddeler, zaman ve uygulayıcılar olarak tanımlanacaktır (Shavelson ve Webb, 1991, s. 10).

1.1.3.4. Çaprazlanmış ve Yuvalanmış Değişkenlik Kaynaklarıyla Desenler

Bir değişkenlik kaynağının (madde) tüm koşullarının diğer varyans kaynağının (birey) tüm koşulları ile gözlemlenmesine çaprazlanmış (*crossed*) desen denir. Bu desende bireyler maddelerle ($b \times m$) çaprazlanmıştır (Shavelson ve Webb, 1991, s. 11). Benzer şekilde her bir öğrenci (b) her bir görevi (g) cevaplandırır ve her bir puanlayıcı (p) da

her bir öğrencinin her görevini puanlarsa bu desen “ $b \times g \times p$ ” şeklinde tümüyle çaprazlanmış bir desen olacaktır.

Çaprazlanmış desen için bütün puanlayıcıların puanlama yapması gerekmektedir. Bazen farklı puanlayıcıların farklı zamanlarda puanlama yapması daha uygun olabilmektedir. Bu gibi durumlarda çaprazlanmış desen değil, yuvalanmış (*nested*) desen kullanılır. Yuvalanmış desen, bir değişkenlik kaynağının (puanlayıcı), sadece bir diğer değişkenlik kaynağını (zaman) gözlemlemesi demektir. Yani farklı öğrencilere (b), farklı görevler sunulursa (g) ve farklı puanlayıcılar (p) farklı öğrenciler tarafından yerine getirilen görevleri puanlarsa, bu desen “ $p : g : b$ ” şeklinde tümüyle yuvalanmış bir desen olacaktır. G kuramı tümüyle çaprazlanmış desenlere ya da çaprazlanmış-yuvalanmış desenlerin karışımına uygulanabilmektedir (Shavelson ve Webb, 1991, s. 11).

Örneğin; bir sosyal psikoloğun utangaçlığın bağlarıyla ilgilendiğini ve utangaçlık düzeyini belirlemek için onlu bir ölçek kullandığını düşünelim. Araştırmada birden fazla puanlayıcı kullanılması durumunda aşağıdaki dört farklı D çalışması deseninden biri kullanılabilir (Crocker ve Algina, 1986, s. 160).

- i. Her öğrenci bir puanlayıcı tarafından puanlanır ve bu puanlayıcı her öğrenciye puan verir.
- ii. Her öğrenci birden fazla puanlayıcı tarafından puanlanır ve her puanlayıcı her öğrenciye puan verir.
- iii. Her öğrenci farklı puanlayıcılar tarafından puanlanır ve her öğrenci için yalnız bir puanlayıcı vardır.
- iv. Her öğrenci birden fazla puanlayıcı tarafından puanlanır ve her öğrenci için farklı puanlayıcılar vardır.

Yukarıdaki dört desende de tek değişkenlik kaynağı *puanlayıcı*dır. İlk iki desende, bütün öğrenciler aynı puanlayıcılar tarafından puanlandığı için ölçmenin değişkenlik kaynağı puanlayıcı, öğrenci ile çaprazlanmıştır. Çaprazlanmış desen, puanlayıcı p ve öğrenci (birey) b ile gösterilmek suretiyle, $p \times b$ şeklinde gösterilir. Çaprazlanmış desenden elde edilen veriye eşleştirilmiş veri denir. Son iki desende ise her öğrenci

farklı puanlayıcılar tarafından puanlandığından puanlayıcılar öğrencilerle yuvalanmıştır. Yuvalanmış desen ise, puanlayıcı p ve öğrenci (birey) b ile gösterilmek suretiyle, $b : p$ şeklinde gösterilir. Yuvalanmış bir desenden elde edilen veriye bağımsız veri denir (Crocker ve Algina, 1986, s. 160).

1.1.3.5. Tesadüfî ve Sabit Değişkenlik Kaynakları

Değişkenlik kaynağının tesadüfî mi sabit mi olacağına karar vermek için öncelikle “Koşulları örneklemin ötesine, evrene genellemek istiyor muyum?” sorusu cevaplandırılmalıdır. Cevap “Evet” ise değişkenlik kaynağı “*tesadüfî*”; “Hayır” ise değişkenlik kaynağı “*sabit*” olarak adlandırılır. Örneklemin boyutu evrenin boyutundan oldukça küçük olduğunda değişkenlik kaynağı tesadüfî kabul edilir (Shavelson ve Webb, 1991, s. 11). Örneğin; araştırmacı 20 maddelik matematik başarı testine uzmanların verdikleri değerlendirmeleri (kesme puanlarını) evrene genellemek istiyorsa değişkenlik kaynağı tesadüfî; evrene genellemek istemiyorsa/sadece çalışmasındaki durumun nasıl olduğunu belirlemek istiyorsa değişkenlik kaynağı sabit olacaktır. Değişkenlik kaynağının tesadüfî veya sabit olması ölçmenin genellenebilirliğini etkilemektedir (Shavelson ve Webb, 1991, s. 12).

1.1.3.6. Genellenebilirlik (G) ve Karar (D) Çalışmaları

Genellenebilirlik (G) çalışmasının amacı, ölçmenin birden çok kullanımını kestirmek ve varyans kaynakları ile ilgili mümkün olan en çok bilgiyi sağlamaktır. G çalışması, mümkün olan en çok değişkenlik kaynağını içerecek biçimde tasarlanmalıdır. G çalışması, kabul edilebilir gözlemlerin evrenini mümkün olan en geniş şekilde tanımlamalıdır (Shavelson ve Webb, 1991, s. 12).

G çalışmasını yürütecek araştırmacı öncelikle hangi örnekleminin evrene genelleneceğine odaklanır. Evren, genellikle gözlenen ölçümlerin örnekleminden daha büyük ölçme durumları seti olarak tanımlanır. Yanıtların zaman içindeki sabitliği, araçların iki ya da daha fazla formundaki puanların eşdeğerliği ya da alt test

puanlarının/ölçekteki maddelerin ilişkileri üzerine bütün çalışmalar G çalışması olarak ele alınabilir (Crocker ve Algina, 1986, s. 158).

Karar (D) çalışması ise, G çalışmasından elde edilen bilgiyi belli bir amaç doğrultusunda, sosyal bilimlerdeki ölçmelerde mümkün olan en iyi deseni oluşturmak için kullanır (Shavelson ve Webb, 1991, s. 12). D çalışması, özel bir amaç için karar vermede kullanılır. D çalışması öğrencileri betimleyen, bir deneyde grupları karşılaştıran veya iki ya da daha fazla değişken arasındaki ilişkileri ortaya çıkaran bilgiler sunar (Crocker ve Algina, 1986, s. 158). D çalışmasında araştırmacı şu aşamaları izler (Shavelson ve Webb, 1991, s. 12):

- i. Genelleme yapılacak evreni tanımlar. (Genelleme yapacağı değişkenlik kaynaklarının sayısını ve genişliğini)
- ii. Ölçme hatalarını ve hata kaynaklarının temelini kesin olarak belirtir. (Görelî ya da mutlak)
- iii. Güvenirliđi maksimum ve hatayı minimum seviyeye getiren alternatif desenlerin etkilerini değerlendirmek için ölçme hatalarının farklı kaynaklarının önemine ilişkin G çalışmasından elde edilen bilgileri kullanır.

G çalışmasının amacı yeterli genellemeyi yapabilecek D çalışmasının planlanmasına yardım etmektir. G çalışması sürecinde puanların varyans kaynakları ve bunlar arasındaki etkileşimler aynı anda ANOVA kullanılarak kestirilir. Kestirilen bu varyans bileşenleri bir sonraki aşama olan D çalışmasında kullanılır. Bu sebeple G çalışması, D çalışmasında kullanılabilen bütün desenleri içerecek biçimde tasarlanmalıdır (Crocker ve Algina, 1986, s. 158).

Bir çalışmanın desenine bakarak o çalışmayı sadece G çalışması veya D çalışması olarak sınıflamak ise mümkün değildir. Bir araştırmacının aynı bölgedeki devlet okullarından ve özel okullardan 2000'er öğrenciyi tesadüfî olarak seçtiđini ve öğrencilere standart bir başarı testi uyguladıđını düşünelim. Eğer amaç testin her iki grup için de eşit seviyede güvenilir olup olmadığını belirlemek ise, bu çalışma G çalışması olarak sınıflandırılır. Diğer bir taraftan araştırmacı madde sayısının (ya da

puanlayıcı sayısının) artması veya azalması durumunda testin güvenilirlik ve geçerliğinin nasıl olacağını D çalışması ile kestirebilir (Crocker ve Algina, 1986, s. 158). Kısacası G kuramı, D çalışmasını G çalışmasından ayırır. D çalışması, G çalışmasından elde edilen bilgileri kullanarak belli bir amaçla yapılan bir ölçmedeki hataları en aza indirmenin yollarını araştırmak için düzenlenir (Brennan 2001, s. 8; Crocker ve Algina 1986, s. 158; Shavelson ve Webb 1991, s. 12).

Bir D çalışmasında, değişkenlik kaynağı sabit veya tesadüfi olarak ele alınabilir. Değişkenlik kaynağı sabit ise, araştırmacı sadece D çalışmasında bulunan durumları genelleme niyetindedir. Değişkenlik kaynağı tesadüfi ise, araştırmacı D çalışmasındaki durumları geniş sayıdaki koşulların bir örnekleme olarak düşünür ve bunu genelleme niyetindedir. Her iki durumda da bütün şartlar altında ölçülerek gözlemlenen ölçümlerin evreni, genelleme evreni (*universe of generalization*) olarak adlandırılır. G çalışması da durumların belirli bir seti altında ölçülerek yürütülür. Bu durumların büyük bir durum setinin temsilcisi olduğu düşünülür. Bu durumlardan alınabilen gözlemlerin evreni, kabul edilebilir gözlemler evreni (*universe of admissible observations*) olarak adlandırılır. G çalışması sonuçları, kabul edilebilir gözlemler evreni araştırmacının genelleme yapacağı evreni içermesi kaydıyla, D çalışmasında araştırmacıya yararlı olacaktır (Crocker ve Algina, 1986, s. 159).

1.1.3.7. Göreceli ve Mutlak Kararlar

Genellenebilirlik kuramında, ölçmenin ne derece genellenebilir olduğu eldeki verilerin karar çalışmasında nasıl kullanıldığına bağlıdır. Sosyal bilimlerdeki ölçmeler genel olarak iki şekilde kullanılır: (a) bireyleri sıralamak (ya da gruplamak) ve (b) bir bireyin (ya da grubun) bilgi, beceri veya tutumunu ya da fikir düzeyini mutlak bir düzeyde belirlemek (Shavelson ve Webb, 1991, s. 13).

Örneğin, özel bir okula öğrenci seçmek amacı ile yapılan 100 soruluk çoktan seçmeli bir öğrenci seçme sınavında “Sınava katılan öğrencilerin test puanları ortalamasının üzerinde puan alanlar başarılı sayılarak okula kayıt yaptırma hakkı elde edeceklerdir” şeklinde okul idaresi tarafından bir açıklama yapılmış olsun (Demirel, 2005, s. 193). Bu

durumda bağıl değerlendirme kullanılacaktır. Bireyin başarısı sınava giren diğer öğrencilerin başarısına bağlıdır. Ölçmenin bu gibi kullanımlarında göreceli kararlar tercih edilmektedir (Shavelson ve Webb, 1991, s. 13).

Ölçümler diğerlerinin ne kadar iyi ya da kötü olduğuna bakılmaksızın bireyin kişisel performansının mutlak düzeyine karar verme amacıyla da kullanılmaktadır. Az önceki örnekte okul idaresinin “Okula kayıt yaptırabilmek için testten en az 90 puan almak gereklidir” şeklinde bir açıklama yaptığını varsayalım (Demirel, 2005, s. 193). Bu durumda ise mutlak değerlendirme kullanılacaktır. Bir bireyin sınavı geçmesi ya da sınavdan kalması diğer bireylerin testte nasıl performans sergilediğine bağlı değil, o bireyin maddeleri doğru cevaplama düzeyine bağlıdır. Kişinin testten aldığı puan, doğru cevapladığı bütün maddelerin yüzdesinin tahminini sağlar. Bu gibi durumlarda ölçüt bağımlı bir sabite değil de kişinin bireysel performansına odaklıdır. Performansın mutlak bir düzeyine dayanan kararlar, mutlak kararlardır (Shavelson ve Webb, 1991, s. 13).

Verilen karar çeşidinin, ölçme hatasının tanımlanması ve hata varyansının büyüklüğü ile ilgisi vardır (Shavelson ve Webb, 1991, s. 13 – 14). Ayrıca göreceli model ile elde edilen genellenebilirlik katsayısı (göreceli genellenebilirlik katsayısı), klasik test kuramındaki güvenilirlik katsayısına denk gelmektedir. Mutlak model ile elde edilen genellenebilirlik katsayısı (mutlak genellenebilirlik katsayısı), göreceli genellenebilirlik katsayısından daha katıdır ve hem ölçülen öğrencilerin sıralamadaki yerlerinin düzenindeki uyumun derecesini hem de ham puanların büyüklüklerindeki uyumun derecesini yansıtır. Mushquash ve O’Connor’a göre (2006) mutlak genellenebilirlik katsayısı, araştırmacı için elde edilen puanların gerçek değerlerinin önemli ya da daha anlamlı olduğunda kullanışlı olabilecek bir katsayıdır (Aktaran: Güler, 2008, s. 28).

1.1.3.8. Genellenebilirlik Katsayısı ve Güvenirlik İndeksi

G kuramı sosyal bilimlerdeki ölçmelerde hataya sebep olan varyansın kaynağına odaklanmasının yanında ayrıca Genellenebilirlik (G) katsayısı denilen bir güvenilirlik katsayısı da ortaya çıkarmaktadır. G katsayısı, göreceli modellerle kullanılır ve gerçek

varyansın, gerçek varyans ile göreceli varyans toplamına bölünmesiyle hesaplanır. G katsayısı, bireyin davranışlarının bir örneklemini gözlenen puanını evren puanına genellemenin ne kadar doğru olduğunu gösterir (Shavelson ve Webb, 1991, s. 14).

KTK'da öğrencinin gerçek puanı, çok sayıdaki paralel ölçümlerin ortalaması olarak tanımlanır. Gerçek puan varyansı bu ortalamaların varyansıdır ve güvenilirlik de gerçek puan varyansının gözlenen puan varyansına (gerçek puan varyansı ve hata puan varyansı) oranıdır. G kuramında ise öğrencinin evren puanı genelleme evrenindeki ölçümlerin ortalaması olarak tanımlanır. Bu ölçümler paralel ölçümler değildir. Genellenebilirlik katsayısı evren puanı varyansının gözlenen puan varyansına oranı olarak tanımlanabilir (Shavelson ve Webb, 1991, s. 14).

Evren puanı ve evren puan varyansı genelleme evrenine bağlıdır. Bu nedenle, aynı D çalışması için farklı genelleme evrenlerini kullanan iki araştırmacı farklı genellenebilirlik katsayıları elde eder (Crocker ve Algina, 1986, s. 159). G kuramında ölçme hataları, göreceli kararlar için ölçme hataları ve mutlak kararlar için ölçme hataları olarak değerlendirilir. G kuramıyla hesaplanan genellenebilirlik katsayısı göreceli ve mutlak değerlendirmeler için ayrı ayrı hesaplanabilir. Kuramda göreceli modeller için G katsayısı elde edildiği gibi, mutlak modeller için de Phi katsayısı elde edilir. Phi gerçek varyansın, gerçek varyans ile mutlak varyansın toplamına bölünmesiyle hesaplanır (Brennan, 2001, s. 13).

Aşağıda bir bireyin (b) bir puanlayıcı (p) tarafından bir zamanda (z) puanlanmasından elde edilen iki değişkenlik kaynaklı çaprazlanmış desene ait puanının (X_{bpz}) bileşenleri gösterilmektedir.

$$\begin{aligned}
X_{bpz} = & \mu && \text{Genel Ortalama} \\
& + \mu b - \mu && \text{Birey Etkisi} \\
& + \mu p - \mu && \text{Puanlayıcı Etkisi} \\
& + \mu z - \mu && \text{Zaman Etkisi} \\
& + \mu bp - \mu b - \mu p + \mu && \text{Birey-Puanlayıcı Etkisi} \\
& + \mu bz - \mu b - \mu z + \mu && \text{Birey-Zaman Etkisi} \\
& + \mu pz - \mu p - \mu z + \mu && \text{Puanlayıcı-Zaman Etkisi} \\
& Xbpz - \mu bp - \mu bz - \mu pz && \text{Artık Etki (residual) (2)} \\
& + \mu b + \mu p + \mu z - \mu && .
\end{aligned}$$

Bu desende gözlenen varyans ise;

$$\sigma^2 (X_{bpz}) = \sigma_b^2 + \sigma_p^2 + \sigma_z^2 + \sigma_{bp}^2 + \sigma_{bz}^2 + \sigma_{pz}^2 + \sigma_{bpz,e}^2 \quad (3)$$

eşitliği ile gösterilir ve bu varyans bileşenlerinin her biri varyans analizi ile kestirilebilir (Shavelson ve Webb, 1991, s. 22 – 23).

Eşitlik 3'ten anlaşıldığı üzere G kuramı ile farklı standart belirleme yöntemlerine göre elde edilen uzman kararlarının varyans bileşenlerinin ve varyans değerlerinin kestirilmesi mümkündür. Daha önce de bahsedildiği gibi alanyazında pek çok standart belirleme yöntemi mevcuttur. Yurtdışındaki çalışmalarla, yöntemlere ait birçok bilginin ortaya konulduğu görülmektedir. Ülkemizde ise standart belirleme yöntemlerinin nasıl sonuçlar verdiğine ve diğer yöntemlerle olan ilişkisine dair fazla çalışma bulunmamaktadır. Özellikle alanyazın incelendiğinde bu araştırma kapsamında yer alan üç yöntemin aynı anda yurtiçi ve yurtdışında fazla çalışılmadığı görülmektedir. Ayrıca yurtiçinde de yurtdışında da farklı yöntemlere göre belirlenen uzman kararlarının G kuramı ile incelenmiş olduğu çalışma sayısı da oldukça azdır. Bu nedenlerle çalışmada aşağıdaki problem cümlesi ve ona ait alt problemlere cevap aranmıştır.

1.2. PROBLEM CÜMLESİ

Angoff, Yes/No ve Ebel standart belirleme yöntemleri arasında nasıl bir ilişki vardır?

1.3. ALT PROBLEMLER

1. Angoff, Yes/No ve Ebel standart belirleme yöntemleri ile belirlenen kesme puanlarına göre başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı bir farklılık var mıdır?
2. Yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflandırılma durumları bakımından uyum var mıdır?
3. Yöntemler arasında uzmanlardan elde edilen minimum geçme puanları (MGP) bakımından uyum var mıdır?
4. Yöntemlere ait uzman kararları arasındaki uyum nasıldır?
5. Yöntemlerin varyans bileşenleri nasıldır?
 - 5.1. Angoff yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?
 - 5.2. Yes/No yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?
 - 5.3. Ebel yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?
 - 5.4. Angoff, Yes/No ve Ebel yöntemleri ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?
6. D çalışmasına göre yöntemlere ait en uygun puanlayıcı sayısı nedir?
 - 6.1. Angoff yönteminde en uygun puanlayıcı sayısı nedir?
 - 6.2. Yes/No yönteminde en uygun puanlayıcı sayısı nedir?
 - 6.3. Ebel yönteminde en uygun puanlayıcı sayısı nedir?

1.4. TANIMLAR VE KISALTMALAR

Kesme Puanı / Geçme Puanı: İstenilen performans düzeyinde olan öğrenci ile yeterli düzeyde olmayan öğrenciyi ayırt edebilen uygun performans noktasını gösterir.

MGP: Bir uzmanın test için belirlediği minimum geçme puanı

KTk: Klasik Test Kuramı

G Kuramı: Genellenebilirlik Kuramı

G Çalışması: Genellenebilirlik Çalışması

D Çalışması: Karar Çalışması

G Katsayısı: Genellenebilirlik Katsayısı

Phi Katsayısı: Güvenirlilik Katsayısı

m: Madde

y: Yöntem

u: Uzman

m x y: Madde – Yöntem Ortak Etkisi

m x u: Madde – Uzman Ortak Etkisi

y x u: Yöntem – Uzman Ortak Etkisi

m x y x u: Madde – Yöntem – Uzman Ortak Etkisi (Artık)

1.5. SINIRLILIKLAR

1. Araştırma geliştirilen ilköğretim 4. sınıf Matematik testi ile sınırlandırılmıştır.

1.6. SAYILTILAR

1. Uzmanların test maddeleri hakkındaki kararlarını, öğrencilerin yalnızca öğrenme düzeylerini dikkate alarak verdikleri varsayılmıştır.

1.7. ARAŞTIRMANIN AMACI VE ÖNEMİ

Bu araştırma kapsamında alanyazında birçok kere incelenmiş iki yöntem olan Angoff ve Yes/No yöntemleri ile kullanılması fazla tercih edilmemiş bir yöntem olan Ebel yöntemi karşılaştırılmıştır. Çalışma kapsamında bu yöntemlere ait kesme puanlarının nasıl olduğu, kesme puanlarının başarılı sayılan öğrenci yüzdesini nasıl etkilediği, kesme puanları arasındaki uyumun ne düzeyde olduğu ve yöntemlere göre uzmanlar arası uyumun derecesi belirlenmeye çalışılmıştır. Ayrıca her bir yöntem için G kuramından

faýdalanılarak yöntemler hakkında detaylı bilgiye ulaşmak ve D çalışması yardımıyla da yöntemler için en uygun uzman sayısını hesaplamak amaçlanmıştır.

Alanyazına bakıldığına çalışmalarda özellikle maddeleri iki aşamalı sınıflamaya tabi tutan ve diğer yöntemlere göre daha ayrıntılı bir süreç gerektiren Ebel yöntemine fazla yer verilmediği görülmektedir. Ülkemizde de Ebel yöntemini konu edinen az sayıda çalışma mevcuttur. Bu açıdan çalışmanın Ebel yöntemine ve diğer iki yöntemle olan ilişkisine dair faydalı bilgiler sunacağı düşünülmektedir. Ayrıca ülkemizde yapılan çalışmalara bakıldığında farklı standart belirleme yöntemlerinin karşılaştırıldığı birkaç araştırmanın olduğu; yöntemlere ait kararların G kuramıyla incelendiği ise oldukça az sayıda araştırmanın olduğu görülmektedir. Ülkemizdeki “Standart Belirleme Yöntemleri” konusuyla ilgili olarak yürütülen çalışmaların azlığı nedeniyle de araştırmanın alanyazına önemli bilgiler sağlayacağı düşünülmektedir.

1.8. İLGİLİ ARAŞTIRMALAR

Andrew ve Hecht (1976) iki test merkezli yöntem olan Ebel ve Nedelsky yöntemlerini iki farklı uzman grubu kullanarak karşılaştırmıştır. İki uzman grubundan her iki yöntem için kararlar alınmış, grupların belirlediği kesme puanlarının Ebel yönteminde de Nedelsky yönteminde de farklılık göstermediği, yani farklı grupların her iki yöntem için de benzer kesme puanı ürettiği sonucuna ulaşılmıştır. Ancak aynı uzman grubunun Ebel yöntemi ve Nedelsky yöntemiyle elde edilen kesme puanları karşılaştırıldığında yöntemlere göre puanlar arasında manidar bir farklılık olduğu görülmüştür.

Brennan ve Lockwood (1980) yaptıkları çalışmada Angoff ve Nedelsky standart belirleme yöntemlerine ait kesme puanlarını Genellenebilirlik kuramını kullanarak karşılaştırmışlar ve kesme puanlarının arasında anlamlı bir farklılık olduğu sonucuna ulaşmışlardır. Bu farklılığın sebepleri olarak da hem bu iki yönteme ait sürecin farklılıkları hem de puanlayıcıların sınır gruba ait minimum yeterlik düzeyini algılama farklılıklarını göstermişlerdir.

Skakun ve Kling (1980) Nedelsky yöntemi ile Ebel yöntemine ait iki varyasyon olan Ebel I ve Ebel II yöntemlerini karşılaştırmışlardır. Maddelerin sadece güçlükleri bakımından sınıflandırılmasını temele alan Ebel I ve maddelerin sadece önem düzeyleri bakımından sınıflandırılmasını temele alan Ebel II yöntemlerine göre kesme puanlarının büyük farklılıklar göstermediği ancak Nedelsky yönteminin bu iki yöntemden oldukça küçük kesme puanı belirlediği sonucuna ulaşmışlardır.

Harasym (1981) Yes/No yöntemi ile Nedelsky yöntemini karşılaştırmak amacıyla 3 paralel testi 2. sınıfta öğrenim gören 3 medikal öğrenci sınıfına uygulamışlar ve analizlerde Rash modelinden faydalanmışlardır. Araştırma sonucunda da iki yönteme ait süreçlerinin, standart belirlemede ve öğrenci başarısı üzerinde manidar etkisi olduğu sonucuna ulaşmışlardır.

Poggio ve diğerleri (1982) tarafından yapılmış bir çalışmada Angoff, Ebel, Nedelsky ve Karşıt gruplar standart belirleme yöntemleri ile elde edilen kesme puanları arasında büyük farklılıklara rastlanmamış ancak en yüksek kesme puanını Ebel yönteminin verdiği, onu da sırasıyla Angoff, Karşıt Gruplar ve Nedelsky yöntemlerinin izlediği sonucuna ulaşılmıştır.

Livingston ve Zieky (1983) çalışmalarında karşıt gruplar yöntemi, sınır grup yöntemi, Angoff ve Nedelsky yöntemlerini okuma ve matematik testleri aracılığıyla karşılaştırmışlardır. Karşıt gruplar ve sınır grup yöntemlerinin birbirine çok yakın kesme puanı ürettiği ve dört yöntem arasında da büyük farklılıkların olmadığı sonucuna ulaşmışlardır.

Mills (1983) çalışmasında Angoff, karşıt grupların iki varyasyonu ve sınır grup yöntemlerini karşılaştırmış, Angoff ve karşıt gruplar (grafik) yöntemlerinin benzer sonuçlar verdiğini; karşıt gruplar yönteminin diğer bir varyasyonu olan QDF yönteminin bazen Angoff ve karşıt gruplar (grafik) ile benzer sonuçlar ürettiğini; sınır grup yönteminin ise diğer tüm yöntemlerden oldukça farklı bir şekilde kesme puanı belirlediğini ortaya koymuştur.

Mills ve Berr (1983) Ebel ve Angoff yöntemlerini, Karşıt Gruplar Yöntemi ile karşılaştırmıştır ve araştırma sonunda Ebel ve Angoff yöntemleri ile elde edilen kesme puanlarının, Karşıt Gruplar yöntemi ile elde edilen kesme puanından daha yüksek olduğu sonucuna ulaşmışlardır.

Appleton (1984) Nedelsky, Ebel ve Karşıt Gruplar standart belirleme yöntemlerini karşılaştırdığı çalışmasında Nedelsky yöntemi ile Karşıt Gruplar yöntemine ait kesme puanlarının birbirine oldukça yakın değerler çıktığı ve Ebel yönteminin daha yüksek bir kesme puanı belirlediği sonucuna ulaşılmıştır.

Bir başka çalışma Norcini ve diğerleri (1987) tarafından yapılmıştır. Çalışmada uzmanlara açıklanmasının – uzmanlar tarafından tamamlanmasının kolay olması ve seçilen uzmanların bu yöntemde deneyimli olmaları nedeniyle Angoff yöntemi kullanılmıştır. Belirlenen 6 uzmandan, 90 soruluk iki paralel test için karar vermeleri istenmiştir. Karar verme süreçleri bir aylık aralarla buluşma öncesi (1. durum), buluşma (2. durum) ve buluşma sonrası (3. durum) olmak üzere üç farklı durumda işlemiştir. Birinci ve ikinci durumda istatistiksel özellikleri benzer paralel testlerden biri; üçüncü durumda ise paralel testlerden diğeri kullanılmıştır. Üç süreçte uzmanlara, uygulama grubunun cevapları, maddelerin doğru cevaplanma yüzdeleri vb. ek bilgiler de sunulmuştur. Çalışmaya başlamadan uzmanlara Angoff yöntemini anlatan bir makale verilmiş ve 10 soruluk küçük bir test ile deneme uygulaması yapmaları sağlanmıştır. Böylece üç süreçte verilen kararlar arasındaki farklılık ortaya konulmaya çalışılmıştır. Birinci süreçte (buluşma öncesi) verilen kararlara ait kesme puanının diğerlerinden daha düşük olduğu; ikinci süreçte (buluşma) verilen kararlara ait kesme puanının ise diğerlerinden daha yüksek olduğu sonucuna ulaşılmıştır. Dolayısıyla buluşma öncesi verilen kararlara göre daha yüksek bir öğrenci yüzdesinin sınavı geçeceği söylenebilir. Ancak üç sürece göre geçen öğrenci yüzdeleri arasında manidar fark bulunamamıştır. Sonuçlara göre buluşma öncesi en tutarsız süreç olurken; buluşma sonrası en tutarlı süreç olmuştur. Ayrıca üç süreç için Genellenebilirlik kuramı ile uzman kararlarının varyans bileşenlerine bakılmıştır. Buluşma öncesinden buluşma sonrasına doğru değişkenlikte azalma görülmüştür. Buna göre uzmanların bir araya geldiklerinde uyumu yakalayabildikleri ve bu uyumu buluşma sonrasında da devam ettirebildikleri

söylenilmektedir. D çalışmasına göre ise sonuçların güvenilirliği için (0,90) buluşma öncesinde 10'dan fazla uzmanın yer alması gerektiği, buluşma sırasında 6 uzman ve buluşma sonrasında da 5 uzmanın yeterli olacağı sonuçları elde edilmiştir.

Chang (1996) araştırmasında Nedelsky ve Angoff Yöntemlerini karşılaştırmış, Nedelsky Yöntemine ait kesme puanının daha düşük olduğunu belirtmiş ancak bu yöntemle hesaplanan kesme puanının öğrenci puanlarına oldukça uygun olduğunu, Angoff yöntemine ait kesme puanının ise öğrenci puanlarından büyük oranda saptığını ortaya koymuştur.

Impara ve Plake (1997) Angoff yöntemi ile Yes/No yöntemini karşılaştırmışlar ve yöntemlerin uygulanabilirliğine dair uzmanlardan görüş almışlardır. Bu amaçla iki farklı çalışma yapmışlardır. Çalışma 1'de, 2. sınıf matematik testi ve 18 uzman; Çalışma 2'de ise 5. sınıf matematik testi ve 20 uzman yer almıştır. Çalışma 1'de uzman grubu rastgele 9 kişilik 2 gruba ayrılmıştır. Gruplardan biri 46 maddelik 2. sınıf matematik testinde Angoff yöntemi ile karar verirken; diğer grup ise aynı test için Yes/No yöntemi ile karar vermiştir. Bu karar verme süreci ise iki farklı durumda gerçekleşmiştir. Birinci durumda uzmanlar kendi kararlarını verirken; ikinci durumda uzmanlara birinci durumda verilen kararlara göre kalan öğrenci yüzdeleri – madde güçlükleri gibi bilgiler de verilerek karar vermeleri sağlanmıştır. İki durum ile elde edilen kesme puanları arasında Angoff yönteminde 8 puanlık ve Yes/No yönteminde de yaklaşık 1 puanlık bir düşüş olduğu görülmüştür. Ayrıca birinci durumdan ikinci duruma standart sapma ve varyansın düştüğü, bir başka deyişle uzmanların kararları arasındaki değişkenliğin azaldığı sonucuna ulaşılmıştır. Angoff yönteminde uzmanlar arası uyum 0,70 iken; Yes/No yönteminde bu uyumun 0,53 olduğu görülmüştür. Uzmanların tümü Yes/No yöntemi ile sadece bir kişi düşünerek karar vermenin daha kolay olduğunu söylemişlerdir. Çalışma 2'de uzman grubu rastgele 10 kişilik 2 gruba ayrılmıştır. Gruplardan biri 89 maddelik 5. sınıf matematik testinde Angoff yöntemi ile karar verirken; diğer grup ise aynı test için Yes/No yöntemi ile karar vermiştir. Bu karar verme süreci de iki farklı durumda gerçekleşmiştir. Birinci durumda uzmanlar kendi kararlarını verirken; ikinci durumda uzmanlara birinci durumda verilen kararlara göre kalan öğrenci yüzdeleri – madde güçlükleri gibi bilgiler de verilerek karar vermeleri

sağlanmıştır. İki durum ile elde edilen kesme puanları arasında Angoff yönteminde yaklaşık 2 puanlık ve Yes/No yönteminde de yaklaşık 1 puanlık bir artış olduğu görülmüştür. Ayrıca birinci durumdan ikinci duruma standart sapma ve varyansın arttığı, bir başka deyişle uzmanların kararları arasındaki değişkenliğin yükseldiği sonucuna ulaşılmıştır. Angoff yönteminde uzmanlar arası uyum 0,90 iken; Yes/No yönteminde bu uyumun 0,56 olduğu görülmüştür. Uzmanların tümü Yes/No yöntemi ile sadece bir kişi düşünerek karar vermenin daha kolay olduğunu belirtmişlerdir.

Hurtz ve Hertz (1999) tarafından yapılan araştırmada 8 farklı test için Angoff yöntemi ile kesme puanları hesaplanmış ve yöntem ile kesme puanı belirlemede en uygun uzman sayısı G kuramı ile belirlenmiştir. Araştırma sonucuna göre Angoff yöntemi ile kesme puanı belirlenirken 10 – 15 uzmanın yeterli olacağı sonucuna ulaşılmıştır.

Buckhendal ve diğerleri (2002) çalışmalarında uzmanlardan 7. Sınıf düzeyindeki bir matematik testi üzerinden iki seferde Angoff ve İşaretleme yöntemlerine uygun şekilde kararlar almışlardır. İlk seferde uzmanların Angoff ve İşaretleme yöntemlerinin normal sürecini izleyerek kararlar vermesini; ikinci seferde ise uzmanların gerçek öğrenci verileri ile (madde güçlükleri ile) kararlar vermesini sağlamışlardır. Bu iki duruma göre kesme puanının ve kalan öğrenci yüzdelerinin nasıl olduğu araştırılmıştır. Angoff yönteminde verilerin uzmanlara verilmesi durumunda kesme puanının küçük oranda düştüğü; İşaretleme yönteminde verilerin uzmanlara verilmesi durumunda ise kesme puanının küçük oranda yükseldiği görülmüştür. Yine de Angoff yönteminde de İşaretleme yönteminde de kesme puanlarında ve kalan öğrenci yüzdelerinde büyük bir farklılığın olmadığı sonucuna ulaşılmıştır. Çalışmada İşaretleme yöntemi ile elde edilen kesme puanının Angoff yöntemi ile elde edilen kesme puanından çok küçük bir farkla daha yüksek olduğu görülmüştür.

Cusimano ve Rothman (2003) Hofstee yöntemini Angoff ve Ebel yöntemleri ile karşılaştırmak amacıyla 96 uzman tarafından OSCE (*Objective Structured Clinical Examination*) için bu üç yönteme göre kararlar almış ve çalışma sonucunda Hofstee yönteminin Angoff ve Ebel yöntemlerine göre daha gerçekçi kesme puanı belirlediği sonucuna ulaşılmıştır.

Tanrıverdi (2006), Nedelsky Yöntemi, Angoff Yöntemi ve Angoff çeşitlemelerinden biri olan Yes/No yöntemlerinin sayısal ve sözel içerikli derslerde geçme puanları üzerinde etkilerinin olup olmadığını araştırmıştır. Bu araştırma sonunda, Türkçe dersi için, farklı standart belirleme yöntemleri geçme puanı üstünde puan alan öğrenci yüzdesine göre ikişerli olarak karşılaştırıldığında, Angoff ve Yes/No yöntemleri arasındaki yüzdeler arasında 0,05 düzeyinde, diğer yöntemler arasında 0,01 düzeyinde fark olduğu görülmüştür. Matematik dersine bakıldığında yöntemlerin ikişerli karşılaştırmalarının tümünde 0,01 düzeyinde farklılaşma görülmüştür. Araştırma ile öğrenciler hakkında verilecek kararlar için kullanılacak standart belirleme yöntemlerinden hangisinin seçileceği öğrenci grubuna, derse, kullanılacak testin yapısına ve öğretmenlere bağlı olarak değişebileceği görülmüştür.

Downing ve diğerleri (2006) çalışmasında sınır grup, karşılaştırmalı gruplar, Angoff, Hofstee ve Ebel yöntemlerinin süreçlerini tartışmışlardır. Sonuç olarak farklı standart belirleme yöntemlerinin farklı standartlar ürettiğini, altın bir standardın olmadığını, tüm standartların politik kararlar olduklarını ve uygulamada da uzman kararlarının öznel kararlar olduğunu belirtmişlerdir.

Taşdelen (2009) araştırmasında Nedelsky yöntemi ile Angoff çeşitlemelerinden biri olan Yes/No yöntemlerini karşılaştırmak için, 2008 yılı ilköğretim 6.sınıf SBS' de yer alan ve 16 maddeden oluşan fen - teknoloji sorularına yönelik 40 kişilik uzman grubuyla kesme puanları belirlemiştir. İki yöntem arasında kesme puanlarına göre geçen öğrenci yüzdeleri arasında 0,05 manidarlık düzeyinde farklılık olduğu ve Nedelsky yöntemine göre geçen öğrenci yüzdesinin Angoff yöntemine göre geçen öğrenci yüzdesine oranla daha büyük olduğu sonucuna ulaşılmıştır. Ayrıca her iki standart belirleme yönteminin birlikte ele alındığı ve tümüyle çaprazlanmış *mpxxy* modeli uygulanarak yapılan G çalışmasında her iki yöntemde de, maddeler arası farklılıkların ortaya çıkarılabildiği ve puanlayıcılar arasında bir tutarlılığın olduğu görülmüştür. Standart belirleme yöntemlerinin ayrı ayrı ele alındığı ve *mxxp* modeli uygulanarak yapılan G çalışmasında ise; Nedelsky yöntemi ile yapılan puanlamada maddeler arası farklılığın ortaya çıkarılabildiği, ancak puanlayıcılar arasında kesme puanı belirlemede uyum olmadığı görülmüştür. Angoff yöntemi ile yapılan puanlamada maddeler arası

farklılığın ortaya çıkarılabildiği ve puanlayıcılar arasında bir tutarlılığın olduğu görülmüştür. G ve Φ katsayılarına bakıldığında ise Angoff ve Nedelsky yöntemleri ile standart belirleme çalışılırken 10 puanlayıcının yeterli olacağı sonucuna varılmıştır.

Ömür ve Selvi (2010) yapmış oldukları araştırmada Angoff, Ebel ve Nedelsky yöntemleriyle ve farklı uzman sayıları kullanılarak elde edilen kesme puanlarının sınıflama tutarlılığı indekslerini incelenmişlerdir. Araştırmanın çalışma grubunda 30 uzman ve 176 ilköğretim 8. sınıf öğrencisi bulunmaktadır. Uzmanların puanlamaları sonucu yukarıda belirtilen yöntemlerle kesme puanları hesaplanmıştır. Araştırmanın sonunda, uzman sayısına bağlı olarak kesme puanlarında önemli bir değişim gözlenmemekle birlikte, üç yöntemde ve üç ayrı uzman sayısında en yüksek sınıflama tutarlılığı indeksinin Nedelsky yöntemiyle elde edildiği görülmektedir.

Çetin (2011) çalışmasında performansa dayanan yöntemlerden işaretleme yöntemini ve test maddelerine dayanan yöntemlerden Angoff yöntemini karşılaştırmıştır. Bu amaçla bir vakıf üniversitesinin hazırlık sınıfı öğrencilerine uygulanan dil sınavına yönelik işaretleme ve Angoff yöntemlerine uygun şekilde 17 uzmandan kararlar alarak kesme puanları belirlemiştir. Bu sınav için mevcut yeterlik puanı ile belirlenen kesme puanlarını karşılaştırmıştır. Angoff yöntemi ile de işaretleme yöntemine ait bir parametrelili ve iki parametrelili tüm lojistik modellerde de mevcut standarttan daha küçük bir kesme puanı elde edildiği görülmüştür. Üniversitenin belirlediği mevcut kesme puanı üzerinde puan alan öğrenci yüzdesi ise işaretleme yöntemine göre belirlenen iki kesme puanından RP50'ye, RP67'ye ve Angoff yöntemine göre geçen öğrenci yüzdeleri arasında 0,01 düzeyinde fark olduğu belirlenmiştir. Cevaplama olasılıkları RP50 ve RP67'ye göre ve işaretleme yöntemi ile Angoff yöntemine göre belirlenen kesme puanları üzerinde puan alan öğrenci yüzdeleri arasında ise önemli bir fark olmadığı saptanmıştır. Ayrıca uzmanlara, geliştirilen standart belirleme sürecini değerlendirme anketi uygulanmış, sorulara verilen cevaplar genel olarak incelendiğinde de uzmanların yöntemleri anlaşılır ve kullanılabilir bulduğu görülmüştür.

Alanyazın incelendiğinde farklı yöntemlerin kullanıldığı ve yöntemlerin birbirleriyle karşılaştırıldığı birçok çalışma olduğu; araştırmalarda ise özellikle Angoff ve Nedelsky yöntemlerinin kullanıldığı görülmektedir. Ebel yöntemi ile elde edilen kesme puanları ile Nedeksly yöntemi ile elde edilen kesme puanları arasında ve Angoff yöntemi ile elde edilen kesme puanları ile Nedeksly yöntemi ile elde edilen kesme puanları arasında manidar bir farklılığın olduğu; bu farklılıkların yöntemlere ait süreçlere ve uzmanların minimum yeterlik düzeylerindeki algılama farklılıklarına dayandırıldığı görülmektedir. Ayrıca Ebel yönteminin farklı varyasyonları (Ebel I ve Ebel II) arasında kesme puanları bakımından bir farklılık olmadığı ancak bu varyasyonların Nedelsky yönteminden oldukça farklı sonuçlar ürettiği görülmektedir.

Uzman sayısı açısından bakıldığında da Angoff yönteminde en az 10 – 15 uzmana ihtiyaç duyulduğu, eğer uzmanlar bir araya gelip karar vereceklerse bu sayının daha az olabileceği sonucuna ulaşılmıştır. Ayrıca uzmanlardan alınan görüşlere göz önünde bulundurulduğunda Angoff yöntemine göre Yes/No yönteminde olduğu gibi sadece bir kişi düşünerek karar vermenin daha kolay olduğu ancak yöntemlere göre uzman kararlarındaki uyuma bakıldığında Angoff yönteminde uzmanların daha yüksek bir uyum yakaladığı söylenebilir. Bunların dışında ise uzmanlara uygulama sonuçlarına ait madde güçlükleri gibi gerçek verilerin verilmesi ve bu şekilde uzmanlardan kararların alınması durumunda kararlar arası değişkenliğin bir çalışmada arttığı başka bir çalışmada ise azaldığı; bir başka deyişle çelişkili sonuçlar elde edildiği görülmektedir.

Bu araştırma kapsamında alanyazında en sık kullanılmış yöntemlerden olan Angoff yöntemi, uzmanlara göre karar vermenin daha kolay olduğu bilgisine ulaşılan Angoff çeşitlemelerinden Yes/No yöntemi ve hakkında oldukça az sayıda araştırma bulunan, iki aşamalı kararlar sonucu kesme puanı belirlemeye olanak sağlayan Ebel yöntemi yer almıştır. Bu üç yöntem ile aynı anda elde edilen kararlar ve kesme puanları arasındaki ilişkinin varlığının belirlendiği ve bu yöntemlere ait kararların G kuramı ile incelendiği bir araştırmaya rastlanmamıştır.

BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın türü, verilerin elde edildiği çalışma grubu, verilerin toplanması – veri toplama araçları ve verilerin analizinde kullanılan istatistiksel yöntemler açıklanmıştır.

2.1. ARAŞTIRMANIN TÜRÜ

Bu araştırmada, üç farklı standart belirleme yöntemi karşılaştırılmıştır. İki ya da daha çok sayıdaki değişken arasında birlikte değişim derecesini belirlemeyi amaçlayan araştırma modelleri ilişkisel türden araştırma modelleridir (Karasar, 1982, s. 85). Çalışma bu açıdan ilişkisel türden araştırma modellerindedir. Betimsel araştırmalar ise, olayların, objelerin, varlıkların, kurumların, grupların ve çeşitli alanların “ne” olduğunu açıklamaya çalışır (Kaptan, 1977, s. 129). Bu araştırma, aynı zamanda her bir yönetime ilişkin betimleyici istatistikleri elde etme, iki ya da daha fazla değişken arasındaki ilişkinin varlığını ve derecesini ortaya koyma yönleriyle betimsel bir çalışmadır. Ayrıca araştırma, geliştirilen bir testin Angoff, Yes/No ve Ebel yöntemlerine dayalı olarak kesme puanlarının belirlenmesi, kesme puanlarının genellenebilirlik kuramı ile incelenmesi ve farklı yöntemlerle elde edilen sonuçların karşılaştırılması yönünden de temel bir araştırmadır. Temel araştırmaların amacı, var olan bilgiye yenilerini katmaktır (Karasar, 1982, s. 24).

2.2. ÇALIŞMA GRUBU

Araştırmada uzmanlar ve öğrenciler olmak üzere iki farklı çalışma grubu yer almaktadır. Araştırmaya katılan, her bir yöntem için kesme puanı belirlemiş olan uzman sayısı 25'tir. Uzmanlardan dördünün, özellikle Ebel yöntemini anlamakta sorun yaşamaları ve diğer dördüne ait kararlardan bazılarının kendi içinde negatif korelasyon vermesi bazılarının da uç değerler olması üzerine bu 8 uzmana ait kararlar analiz dışında tutulmuştur. Bu sebeple uzman grubu 17 kişiden oluşmaktadır. Uzmanlara elverişli (amaca uygun) örnekleme yöntemiyle ulaşılmıştır ve uzman seçiminde

uzmanların gerekli bilgi beceriye sahip olmalarının yanında gönüllülük durumları esas alınmıştır. Üç standart belirleme sürecinde de aynı uzman grubu yer almaktadır. Yani bütün uzmanlardan üç yönteme göre tüm maddeler için kararlar alınmıştır. Uzman grubu, alanında deneyimli sınıf öğretmenlerinden oluşmaktadır.

Ayrıca Tablo 7’de de görüldüğü üzere araştırmaya toplam 489 ilköğretim 4. sınıf öğrencisi katılmıştır. Bu öğrencilere araştırmacı tarafından geliştirilmiş 20 soruluk matematik başarı testi uygulanmıştır.

Tablo 7: Araştırmaya Katılan Okullar ve Öğrenci Sayıları

Araştırmaya Katılan Okullar	Öğrenci Sayısı
Batıkent İlköğretim Okulu	122
Abay İlköğretim Okulu	116
Afşin Bey İlköğretim Okulu	83
Prof. Dr. Mehmet Sağlam İlköğretim Okulu	68
Kurtuluş İlköğretim Okulu	67
Namık Kemal İlköğretim Okulu	33
TOPLAM	489

2.3. VERİLERİN TOPLANMASI VE VERİ TOPLAMA ARAÇLARI

Araştırma verilerinin toplandığı gruplar, öğrenciler ve uzmanlardır. Öğrenciler ve uzmanlardan aynı test aracılığıyla veri toplanmıştır. Araştırmada kullanılan öğrenci verileri, ilköğretim 4. sınıf matematik dersi “Bölme ve Kesirler” alt öğrenme alanlarına yönelik araştırmacı tarafından hazırlanmış 20 soruluk bir başarı testinden (*EK 1*) alınan puanlardır. Araştırmada kullanılan diğer verileri oluşturan, uzmanların hazırlanan bu başarı testindeki maddeler için üç farklı standart belirleme yöntemine göre vermiş oldukları kararlar ise, yine araştırmacı tarafından hazırlanmış “Uzman Görüşleri İçin Araştırma Formu” (*EK 2*) ile 17 uzmandan elde edilmiştir.

Başarı testinin geliştirilmesi aşamasında, “Bölme ve Kesirler” alt öğrenme alanlarından seçilmiş 11 kazanıma yönelik 40 soruluk bir test oluşturulmuştur. Ön uygulama olarak

bu test 2010 öğretim yılının Aralık ayında 272 ilköğretim 5. sınıf öğrencisine yaklaşık bir ders saati içinde uygulanmıştır. Ön uygulama sonucunda veriler Excel'e aktarılarak ve ITEMAN programı kullanılarak analiz edilmiştir. Madde seçiminde madde ayırt edicilik gücü indeksleri (r_{jx}) ve madde güçlük indeksleri (P_j) dikkate alınmıştır. Tablo 8'de ön uygulama sonuçlarına göre madde istatistikleri özetlenmiştir.

Tablo 8: Ön Uygulama Sonuçlarına Göre Madde İstatistikleri (n=272)

M	P_j	S_j	r_{jx}	M	P_j	S_j	r_{jx}
1	0,92	0,27	0,59	21	0,96	0,21	0,58
2	0,91	0,29	0,43	22	0,64	0,48	0,66
3	0,60	0,49	0,62	23	0,68	0,47	0,74
4	0,75	0,44	0,81	24	0,88	0,32	0,62
5	0,97	0,17	0,30	25	0,21	0,41	0,42
6	0,89	0,31	0,67	26	0,78	0,41	0,77
7	0,74	0,44	0,45	27	0,71	0,45	0,49
8	0,82	0,38	0,64	28	0,86	0,34	0,53
9	0,82	0,38	0,59	29	0,80	0,40	0,41
10	0,98	0,13	1,00	30	0,60	0,49	0,52
11	0,85	0,35	0,68	31	0,81	0,40	0,57
12	0,78	0,41	0,45	32	0,57	0,49	0,55
13	0,89	0,32	0,85	33	0,74	0,44	0,82
14	0,39	0,49	0,68	34	0,93	0,26	0,87
15	0,64	0,48	0,62	35	0,90	0,29	0,79
16	0,88	0,33	0,57	36	0,93	0,25	0,53
17	0,46	0,50	0,56	37	0,90	0,30	0,83
18	0,67	0,47	0,50	38	0,46	0,50	0,57
19	0,76	0,43	0,80	39	0,82	0,38	0,63
20	0,84	0,37	0,57	40	0,89	0,32	0,65

M: Madde; P_j : Madde güçlüğü; S_j : Madde Standart Sapması; r_{jx} : Çift Serili Korelasyon Katsayısı

Tablo 8 incelendiğinde madde ayırt edicilik gücü indeksleri 0,30 ile 1,00 arasında; madde güçlük indeksleri ise 0,21 ile 0,98 arasında değiştiği görülmektedir. Analiz sonuçlarına göre madde ayırt edicilik gücü indeksi 0,30 ve üstünde olan 40 madde de nihai forma alınabilmektedir. Bu durumda çeldiricilerin durumları göz önünde

bulundurularak aynı kazanımı ölçmeye yönelik hazırlanmış maddeler arasından çeldiricileri en iyi işleyen ve diğerlerine göre daha yüksek ayırt ediciliğe sahip olan maddeler seçilmiştir. Herbir kazanım için en az bir madde nihai forma alınacak şekilde madde güçlük indeksi 0,50 civarında olan maddeler ile daha kolay ve daha zor maddeler seçilmiştir.

Sonuç olarak nihai forma alınan maddeler 3, 4, 8, 9, 12, 13, 14, 15, 17, 19, 20, 22, 23, 26, 27, 30, 31, 33, 37, 39 numaralı maddelerdir. Seçilen maddelerden oluşturulan nihai teste ait özellikler ön uygulama sonuçlarından kestirilmiş olup Tablo 9’da yer almaktadır.

Tablo 9: Nihai Testin Kestirilen Özellikleri

	K	n	S _j	\bar{X}	\bar{P}	KR-20
Nihai Test	20	272	4,24	14,41	0,72	0,84

Tablo 9’da görüldüğü gibi ön uygulama sonuçlarına göre oluşturulan testin güvenilirliği için KR-20 iç tutarlılık güvenirlik katsayısı 0,84 olarak kestirilmiştir. Murphy ve Davidshofer’a göre (1991, s. 103) sınıf içi başarı testlerinde iç tutarlılık katsayısının en az 0,75 olması gerekmektedir. Bu bakımdan kestirilen değerinin yeterli olduğu söylenebilir.

Testin kapsam geçerliğini sağlamak amacıyla da 4 alan uzmanından görüş alınmıştır. Uzman kararların uyumunu belirlemek amacıyla Cohen’in Kappa istatistiği 0,82 olarak hesaplanmıştır. Bu değer uzmanlar arasında neredeyse mükemmel uyuma olduğunu göstermekte ve geliştirilen testin kapsam geçerliğinin yüksek olduğu anlamına gelmektedir.

2.3.1. Öğrencilerden Verilerin Toplanması

Araştırmacı tarafından hazırlanan 20 soruluk matematik dersi “Bölme ve Kesirler” alt öğrenme alanlarına yönelik başarı testi 489 ilköğretim 4. sınıf öğrencisine 35 dakika süreyle uygulanmıştır. Uygulamada öğrencilere puanların, okulun en iyi sınıfını

belirlemede kullanılacağı söylenerek öğrenciler motive edilmiştir. Tüm uygulamalarda sınıf öğretmeni ile birlikte araştırmacı da sınıfta bulunmuştur. Nihai uygulama sonuçlarına göre testin özellikleri Tablo 10’da yer almaktadır.

Tablo 10: Nihai Uygulama Sonuçlarına Göre Test İstatistikleri

	K	n	S_j	\bar{X}	\bar{P}	KR-20
Nihai Test	20	489	4,08	12,58	0,63	0,82

Tablo 10’a göre testin nihai uygulama sonuçlarından elde edilen iç tutarlılık güvenirlik katsayısı 0,82’dir. Nihai testin geçerlik çalışması için asıl uygulamanın yapıldığı gruptan ulaşılabilen 180 öğrencinin 2010 – 2011 birinci dönem karne notları ile uygulama sonucunda aldıkları puanlar arasındaki korelasyona bakılmıştır. Bu iki ölçüme ait puan aralıklarının farklı olması sebebiyle Spearman Brown Sıra Farkları Korelasyon Katsayısı hesaplanmış olup bu değer 0,62 ($p < 0,01$) olduğu görülmüştür. Bu durum uygulama sonuçları ile karne notları arasında 0,01 hata ile yani %99 olasılıkla orta düzeyde anlamlı bir ilişki olduğunu göstermektedir.

2.3.2. Uzmanlardan Verilerin Toplanması

Nihai testte yer alan 20 soru için 25 uzmandan üç standart belirleme yöntemine göre kararlar alınmıştır. Uzmanlardan dördünün, özellikle Ebel yöntemini anlamakta sorun yaşamaları ve diğer dördüne ait kararların da bazılarının kendi içinde negatif korelasyon vermesi bazılarının da uç değerler olması üzerine bu 8 uzmana ait kararlar analiz dışında tutulmuştur. Uzmanlardan karar alınan bu üç yöntemde sürecin nasıl işlediği aşağıda kısaca açıklanmıştır:

- Angoff Yöntemi için uzmanlardan, geçme – kalma sınırındaki (minimum yeterlik düzeyindeki) 100 öğrenciden kaçının o test maddesini doğru cevaplayabileceğini düşünmeleri istenmiştir. Bu şekilde 20 soru için bir doğru cevaplanma yüzdesi vermeleri sağlanmıştır. Her uzmana ait yüzdeler toplanıp soru sayısı olan 20’ye bölünerek uzmanlara ait minimum geçme puanı (MGP)

belirlenmiştir. Yönteme ait kesme puanı ise MGP'lerin ortalaması alınarak hesaplanmıştır.

- Yes/No Yöntemi için de uzmanlardan, yine geçme – kalma sınırındaki (minimum yeterli düzeyindeki) sadece 1 öğrenciyi düşünmeleri ve o öğrencinin soruları doğru cevaplayıp cevaplayamama durumlarından uygun olanını işaretlemeleri istenmiştir. Uzman kararlarına göre öğrencinin doğru cevaplayabileceği sorulara “1” puan; doğru cevaplayamayacağı sorulara ise “0” (sıfır) puan verilerek uzmanlara ait MGP'ler hesaplanmıştır. Yönteme ait kesme puanı da MGP'lerin ortalamaları alınarak elde edilmiştir.
- Ebel Yöntemi için ise diğer iki yöntemden daha farklı olarak uzmanlardan öncelikle bütün sorular için “Zorluk” ve “Uygunluk” olmak üzere ikili bir sınıflandırma yapmaları istenmiştir. Her bir soru için *Kolay-Orta-Zor* ve *Gerekli-Önemli-Kabul Edilebilir-Tartışılabilir* seçeneklerinden uygun olanına soru numaralarını yazmaları istenmiştir. Bütün sorular bu iki düzeye yerleştirildikten sonra her hücrede bulunan soru/sorular için yine geçme – kalma sınırındaki (minimum yeterli düzeyindeki) 100 öğrenciden kaçının o test maddesini/maddelerini doğru cevaplayabileceğini düşünmeleri istenmiştir ve hücredeki sorular için ortak bir yüzde yazmaları sağlanmıştır. Elde edilen veriler Excel'de hücrelerdeki soru sayısı ve yüzdeler çarpımları toplanarak uzmanlara ait MGP'ler hesaplanmıştır. Yönteme ait kesme puanı da yine MGP'lerin ortalamaları alınarak elde edilmiştir.

2.4. VERİLERİN ANALİZİ

Araştırmanın her bir alt probleminin çözümlenmesinde kullanılan istatistiksel teknikler aşağıda sırasıyla verilmiştir.

Birinci alt problemin çözümünde; bağımlı iki oran arasındaki farkın testi yapılmıştır. Bağımlı iki oran arasındaki farkın testi, yöntemlere göre başarılı ve başarısız sayılan

öğrenci sayılarına ve yüzdelerine ait 2×2 'lik bir tablo yardımıyla hesaplanabilmektedir. Aşağıda Tablo 11'de bu 2×2 'lik tablolara ve değerlerin anlamlarına yer verilmiştir.

Tablo 11: Frekans ve Oranlara Göre İstatistiksel Önem Testi

		Yöntem 2					Yöntem 2		
		Kaldı	Geçti				Kaldı	Geçti	
Yöntem 1	Geçti	A	B	A + B	Yöntem 1	Geçti	a	b	p1
	Kaldı	C	D	C + D		Kaldı	c	d	q1
		A + C	B + D	N			q2	p2	1.00

Yukarıda frekanslara ait tabloda Yöntem 1 ve Yöntem 2'ye göre geçen – kalan öğrenci sayıları yer almaktadır. B, her iki yöntemden de geçen öğrenci sayısını; C her iki yöntemden de kalan öğrenci sayısını; A birinci yöntemden geçip ikinci yöntemden kalan öğrenci sayısını; D ise ikinci yöntemden geçip birinci yöntemden kalan öğrenci sayısını göstermektedir.

Oranlara ait tabloda Yöntem 1 ve Yöntem 2'ye göre geçen – kalan öğrenci sayılarının N' e bölümü olan geçen – kalan öğrenci oranları yer almaktadır. a, birinci yöntemden geçip ikinci yöntemden kalan öğrenci oranını; b her iki yöntemden de geçen öğrenci oranını; c her iki yöntemden de kalan öğrenci oranını; d ise ikinci yöntemden geçip birinci yöntemden kalan öğrenci oranını göstermektedir. Bu durumda p_1 , a + b'ye; p_2 ise b + d'ye eşittir.

z istatistiği her iki yöntemden de geçen öğrenci oranları arasındaki farkın oranlara ait standart sapmaya bölümüyle elde edilir. Bu durumda standart sapma;

$$S_{p_1-p_2} = \sqrt{\frac{a+d}{N}}$$

olarak ifade edilir (4) ve z istatistiği de,

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{a+d}{N}}} \text{ şeklinde hesaplanır (Ferguson, 1971, s. 162 – 164). (5)}$$

489 öğrenciye ait veriler doğru cevaplar 1 ve yanlış - boş cevaplar 0 kodlanarak Excel programına aktarılmıştır. Böylece öğrencilere ait doğru cevap sayıları ve öğrencilerin testten aldıkları puanlar belirlenmiştir. Ardından üç yöntem için alınmış olan uzman kararları Excel'e girilerek her bir uzmana ait MGP belirlenmiştir. MGP'lerin ortalaması alınarak da yöntemler için kesme puanları elde edilmiştir. MGP'ler de kesme puanları da daha anlamlı olmaları adına 100 üzerinden değerlendirmeye alınacak şekilde, yani 5 ile çarpılarak hesaplanmıştır. Kesme puanlarına göre her bir yöntem için başarılı sayılan öğrenci sayıları ve yüzdeleri belirlenmiştir. Yöntemler ikili olarak sırasıyla Angoff – Yes/No, Ebel – Angoff ve Ebel – Yes/No olmak üzere karşılaştırılarak Excel'de üç farklı z istatistiği hesaplanmıştır. Bu şekilde yöntemlere göre geçen öğrenci yüzdeleri arasında 0,01 hata ile manidar bir fark olup olmadığına bakılmıştır.

İkinci alt probleme cevap verebilmek amacıyla öğrenci puanları üç yöntemle elde edilen kesme puanlarına göre 1-0 kodlanarak yapay ikili hâle getirilip yöntemlerin ikili uyumlarına kategorik verilerin uyumunu ölçen Cohen'in Kappa istatistiği ile bakılmıştır. Cohen (1960) tarafından geliştirilen Kappa istatistiği, iki veya daha fazla gözlemcinin yaptığı değerlendirmeler arasındaki uyuşmayı belirlemek için kullanılır. Bu uyum -1 ile +1 arasında değer alır. Sıfır değeri tesadüfi uyuşmayı, negatif değerler tesadüfi olmaktan daha kötü bir uyuşmayı ve +1 değeri ise mükemmel uyuşmayı temsil eder. Uyuşma oranlarının yüksekliği veya düşüklüğüne ilişkin standartlar farklı bilim insanları tarafından değişik şekillerde açıklanmıştır. Kappa sayısı 0,40 ile 0,75 arasında değer aldığındaki makul bir uyuşma ve 0,75'ten büyük bir değer aldığındaki da mükemmel uyuşma anlamına gelir (Şencan, 2005, s. 265 – 268). Landis ve Koch'a göre (1977) ise Kappa istatistiği 0,00 – 0,20 arasında düşük; 0,21 – 0,40 arasında makul; 0,41 – 0,60 arasında orta; 0,61 – 0,80 arasında önemli ve 0,81 – 1,00 arasında mükemmel uyumu göstermektedir (Aktaran: Şencan, 2005, s. 267 – 268).

Üçüncü alt probleme cevap verebilmek adına her bir yöntem için uzman kararlarından elde edilen 17 adet MGP arasındaki ilişkiye, MGP'lerin yapısını da dikkate alarak

Pearson Momentler Çarpımı Korelasyon Katsayısı ile ve kesme puanlarının ortalamaları arasındaki farka da bağımlı gruplar t testi ile bakılmıştır.

Dördüncü alt problem olan yöntemlerde uzmanların uyumunun nasıl olduğunu belirlemek amacıyla Angoff, Yes/No ve Ebel yöntemlerinde uzmanların maddelere vermiş oldukları kararlar arasındaki ilişkiye bakılmıştır. Her bir yöntemde karar vermiş olan 17 kişilik uzman grubunun, her bir test maddesi için verdikleri kararların uyumuna Angoff ve Ebel yöntemleri için Kendall'ın uyuşma katsayısı (W) ile bakılırken; Yes/No yöntemi için kategorik verilerin uyumunu belirlemeye yönelik olan Cohen'in Kappa istatistiğinden faydalanılmıştır. Kendall W istatistiği, ikiden fazla puanlayıcının kullanılması durumunda, puanlayıcılar arası güvenilirliğin belirlenmesinde kullanılan parametrik olmayan istatistiksel bir tekniktir (Howell, 2007, s. 290 – 292). Eğer ikiden daha çok sayıda uzman sıralı ölçek verisinde puanlama yaptıysa ve bu çoklu sayıda uzman için tek bir uyum katsayısı belirlenmek istenirse Kendall'ın W katsayısından yararlanılmaktadır. Kendall'ın W katsayısı 0 (hiç uyuşma olmaması) ile 1 (tam olarak uyuşma olması) arasında değer alır. Szymanski ve diğerleri'ne göre (1993) yapılan değerlendirmelerde puanlayıcılar arasındaki uyumun en az 0,80 düzeyinde olması gerekmektedir (Aktaran: Deliceoğlu, 2009, s. 40).

Beşinci alt problemin ilk üç problemine cevap verebilmek amacıyla uzmanların Angoff, Yes/No ve Ebel yöntemlerinde maddelere vermiş oldukları kararlara ait varyans bileşenleri ve varyans yüzdeleri EduG (2010) programıyla her bir yöntem için ayrı ayrı tek değişkenli çaprazlanmış modelle ($m \times u$) belirlenmiştir. Model aşağıda Tablo 12'de yer almaktadır.

Tablo 12: Tek Değişkenli Çaprazlanmış Modelle Varyans Bileşenlerinin Kestirilmesi

Değişkenlik Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans
Madde (m)	SS_m	$n_m - 1$	$MS_m = SS_m / n_m - 1$	$\sigma^2_{(m)}$
Uzman (u)	SS_u	$n_u - 1$	$MS_u = SS_u / n_u - 1$	$\sigma^2_{(u)}$
m x u, e	SS_{mu}	$(n_m - 1)(n_u - 1)$	$MS_{mu} = SS_{mu} / n_{mu} - 1$	$\sigma^2_{(mu)}$

Beşinci alt problemin son alt problemine cevap verebilmek amacıyla ayrıca uzmanların Angoff, Yes/No ve Ebel yöntemlerinde maddelere vermiş oldukları kararlara ait varyans bileşenleri ve varyans yüzdeleri EduG (2010) programıyla yöntemlerin üçünü de kapsayacak şekilde iki değişkenli çaprazlanmış modelle ($m \times y \times u$) belirlenmiştir. Model aşağıdaki Tablo 13’de yer almaktadır.

Tablo 13: İki Değişkenli Çaprazlanmış Modelle Varyans Bileşenlerinin Kestirilmesi

Değişkenlik Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans
Madde (m)	SS_m	$n_m - 1$	$MS_m = SS_m / n_m - 1$	$\sigma^2_{(m)}$
Yöntem (y)	SS_y	$n_y - 1$	$MS_y = SS_y / n_y - 1$	$\sigma^2_{(y)}$
Uzman (u)	SS_u	$n_u - 1$	$MS_u = SS_u / n_u - 1$	$\sigma^2_{(u)}$
m x y	SS_{my}	$(n_m - 1)(n_y - 1)$	$MS_{my} = SS_{my} / n_{my} - 1$	$\sigma^2_{(my)}$
m x u	SS_{mu}	$(n_m - 1)(n_u - 1)$	$MS_{mu} = SS_{mu} / n_{mu} - 1$	$\sigma^2_{(mu)}$
y x u	SS_{yu}	$(n_y - 1)(n_u - 1)$	$MS_{yu} = SS_{yu} / n_{yu} - 1$	$\sigma^2_{(yu)}$
m x y x u, e	$SS_{myu,e}$	$(n_m - 1)(n_y - 1)(n_u - 1)$	$MS_{myu} = SS_{myu,e} / n_{myu} - 1$	$\sigma^2_{(myu)}$

Son olarak altıncı alt problem için ise yöntemlere ait uzman kararları üzerinde EduG programından faydalanarak D çalışması yapılmış, her bir yöntem için farklı uzman sayısı olması durumunda uzman kararlarına ait G ve Phi katsayıları hesaplanmıştır. Uzman sayılarının 5’er azaltılması ve 5’er artırılması durumunda (7, 12, 22, 27 ve 32 uzman olması durumunda) yöntemlerdeki G ve Phi katsayıları arasındaki değişime bakılmış ve her bir yöntem için en uygun uzman sayısı belirlenmiştir.

BÖLÜM III

BULGULAR VE YORUMLAR

Bu bölümde sırasıyla her bir alt problem için elde edilen bulgular ve bulgulara ilişkin yorumlar yer almaktadır.

3.1. Alt Problem 1: Angoff, Yes/No ve Ebel standart belirleme yöntemleri ile belirlenen kesme puanlarına göre geçen öğrenci yüzdeleri arasında anlamlı bir farklılık var mıdır?

Bu alt problemin çözümü için 17 uzmandan her bir yöntem için belirlenen yönergelere göre 20 test maddesi hakkında karar vermeleri istenmiştir. Angoff yöntemi, Yes/No yöntemi ve Ebel yöntemi için verilen kararlar sırasıyla Ek 3, 4, 5'te verilmiştir. Her bir uzmana ait bu kararların ortalaması ile uzmanlara ait minimum geçme puanları (MGP) hesaplanmıştır. Teste ait kesme puanı ise MGP'lerin ortalamaları alınarak elde edilmiştir. Her bir yöntem için uzmanlara ait MGP'ler ve hesaplanan kesme puanları Tablo 14'te gösterilmektedir.

Tablo 14: Yöntemler İçin Elde Edilen MGP'ler

Uzman	Angoff Y.	Yes/No Y.	Ebel Y.	Uzman	Angoff Y.	Yes/No Y.	Ebel Y.
1	54,5	55	53,5	10	35,75	45	33,25
2	61,5	50	48	11	47,25	65	46
3	62	65	60,5	12	49,5	55	48,5
4	48,5	60	19	13	58	45	58,25
5	48	40	41,5	14	29	55	21,75
6	66	80	68	15	59,5	40	58,25
7	46,65	55	47	16	65,5	75	67,75
8	65,35	75	49	17	53,25	90	57,75
9	24,75	35	23				
Kesme Puanı					51,47	57,94	47,12

Tablo 14’te de görüldüğü gibi uzmanların maddelere verdikleri kararların farklılığı uzmanlara ait MGP’lerin de farklı olmasına neden olmuştur. Bu MGP’lerin ortalaması olan teste ait nihai kesme puanları da yöntemden yönteme farklılık göstermektedir. Tablo 14’e göre en yüksek kesme puanı 57,94 ile Yes/No yönteminde elde edilmiştir. Yes/No yöntemini 51,47 kesme puanı ile Angoff yöntemi takip etmektedir. En düşük kesme puanı ise Ebel yöntemi (47,12) ile belirlenmiştir. Tablo 14’te de görülen yöntemlere göre elde edilen kesme puanlarındaki bu farklılık başarılı kabul edilen öğrenci sayılarını ve yüzdelerini de etkilemiştir. Yöntemlere ait kesme puanlarına göre başarılı kabul edilen öğrenci sayıları ve yüzdeleri hesaplanmış olup aşağıda Tablo 15’te verilmiştir.

Tablo 15: Yöntemlere Ait Kesme Puanları, Başarılı Kabul Edilen Öğrenci Sayıları ve Yüzdeleri (n=489)

Yöntem	Kesme Puanı	Başarılı Kabul Edilen Öğrenci Sayısı	Başarılı Kabul Edilen Öğrenci Yüzdesi
Angoff Y.	51,47	327	0,67
Yes/No Y.	57,94	293	0,60
Ebel Y.	47,12	371	0,76

Tablo 15’te de görüldüğü gibi yöntemlere ait belirlenen kesme puanlarına göre başarılı kabul edilen öğrenci sayıları ve yüzdeleri farklılık göstermektedir. En düşük kesme puanının ait olduğu Ebel yöntemi, diğer yöntemlere göre başarılı kabul edilen öğrenci sayısı ve yüzdesinin en fazla olduğu yöntemdir. Ebel yöntemini Angoff yöntemi takip etmektedir. En az başarılı kabul edilen öğrenci sayısı ve yüzdesi ise en yüksek kesme puanının (57,94) elde edildiği Yes/No yöntemine aittir. Yöntemlere ait başarılı kabul edilen öğrenci yüzdeleri arasındaki bu farklılıkların anlamlılığı bağımlı iki oran arasındaki farkın testi ile sınanmış olup sonuçları Tablo 16’da verilmiştir.

Tablo 16 incelendiğinde yöntemlere ait kesme puanlarına göre geçen öğrenci yüzdeleri ikili olarak karşılaştırıldığında tüm yüzdeler arasında 0,01 manidarlık düzeyinde, bir başka deyişle %99 olasılıkla anlamlı bir farklılık olduğu görülmektedir. Angoff ve Yes/No yöntemleri için geçen öğrenci yüzdeleri açısından elde edilen bu sonuç Tanrıverdi’nin çalışması (2006) ile de örtüşmektedir.

Tablo 16: Kesme Puanları, Başarılı Sayılan Öğrenci Sayıları-Yüzdeleri ve z değerleri

Karşılaştırılan Yöntemler	Kesme Puanı	Başarılı Sayılan Öğrenci Sayısı	Başarılı Sayılan Öğrenci Oranı	z
Angoff Yöntemi	51,47	327	0,67	5,85**
Yes/No Yöntemi	57,94	293	0,60	
Ebel Yöntemi	47,12	371	0,76	6,63**
Angoff Yöntemi	51,47	327	0,67	
Ebel Yöntemi	47,12	371	0,76	8,85**
Yes/No Yöntemi	57,94	293	0,60	

**p < 0,01

Başarılı sayılan öğrenci yüzdelerindeki bu manidar farklılıkların nedeni üç yöntemde de verilen kararların farklı dayanakları olmasına bağlanabilir. Angoff yönteminde uzmanlardan 100 kişilik geçme-kalma sınırındaki öğrenciden kaçının maddeyi doğru cevaplayabileceği düşünceleri istenirken; Yes/No yönteminde uzmanlardan geçme-kalma sınırındaki sadece bir öğrenciyi dikkate almaları istenmektedir. Bu kararlar doğrultusunda da MGP'lerin ortalaması alınarak kesme puanları hesaplanmaktadır. Ebel yönteminde ise bu iki yöntemden tamamen farklı olarak uzmanlardan maddeleri önce Zorluk ve Uygunluk düzeylerine göre 3×4 ' lük bir tabloya yerleştirmeleri, sonrasında da her hücrede bulunan soru/sorular için geçme-kalma sınırındaki 100 öğrenciden kaçının soruyu doğru cevaplayabileceği düşünceleri istenmektedir. Hücrelerdeki madde sayısı ile bu yüzdeler çarpılarak MGP'ler hesaplanmakta ve kesme puanı da MGP'lerin ortalaması olmaktadır. Böylece yöntemlerin süreçlerindeki farklılık kesme puanlarına yansımaktadır. Kesme puanlarındaki farklılık da geçen öğrenci yüzdelerini etkilemektedir. Ayrıca Impara ve Plake (1997) çalışmalarında uzmanların Yes/No yönteminde daha kolay karar verdiği sonucuna ulaşmışlardır. Uzmanlar bu çalışmada Angoff yönteminde karar vermede 100 kişilik bir grubu düşünmenin daha zor olduğunu belirtmişlerdir. Uzmanların yöntemlere ait süreçlerdeki bu tip algıları da kararları etkileyebilir. Dolayısıyla kesme puanları ve başarılı sayılan öğrenci sayıları-yüzdeleri de bu durumdan etkilenebilir.

3.2. Alt Problem 2: Yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflanma durumları bakımından uyum var mıdır?

Bu alt problemin çözümü için gerçekte sürekli olan öğrenci puanları, elde edilen kesme puanları temel alınarak yapay ikili hâle getirilmiştir. Bu amaçla öğrenci puanları üç yönteme göre 1-0 puanlanmış ve yöntemler arasındaki ikili uyumlara Cohen'in Kappa istatistiği ile bakılmıştır. Kappa istatistiği ile beraber Pearson ve Spearman korelasyon katsayıları da Tablo 17'de yer almaktadır.

Tablo 17: Yöntemler arasındaki uyum için hesaplanan korelasyon katsayıları

Yöntemler	Kappa	Spearman	Pearson
Angoff – Yes/No	0,85**	0,86**	0,86**
Angoff – Ebel	0,78**	0,81**	0,81**
Yes/No – Ebel	0,64**	0,69**	0,69**

** p < 0,01

Tablo 17 incelendiğinde tüm yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflanma durumları bakımından pozitif yönde ve 0,01 manidarlık düzeyinde uyum olduğu görülmektedir. Angoff ve Yes/No yöntemleri ile Angoff ve Ebel yöntemleri arasında öğrencilerin başarılı-başarısız olarak sınıflanma durumları bakımından mükemmel bir uyum görülürken; Yes/No ve Ebel yöntemleri arasında makul bir uyum söz konusudur.

Angoff ve Yes/No yöntemlerinin arasındaki bu mükemmel uyumun nedeni iki yöntemde de uzmanların sadece sınır grup öğrencisine odaklanıyor olmasından ileri gelebilir. Uzmanlar bu iki yöntemde başka bir görevle yükümlü değildirler. Ayrıca Yes/No yöntemi daha önce de bahsedildiği gibi Angoff yönteminden türetilmiştir. Angoff ve Ebel yöntemleri arasındaki uyum ise yöntemlerde uzmanlardan sınır gruba dair yüzde alınıyor olması nedeniyle yüksek olabilir. Yes/No ve Ebel yöntemleri arasındaki makul ancak diğerlerine kıyasla daha düşük bu uyumun nedeni iki yöntemdeki sürecin birbirinden oldukça farklı işliyor olması olabilir.

3.3. Alt Problem 3: Yöntemler arasında uzmanlardan elde edilen minimum geçme puanları (MGP) bakımından uyum var mıdır?

Bu alt problemin çözümü için öncelikle yöntemlere ait MGP'lerin normal dağılıma sahip olup olmadığına bakılmıştır. Bu amaçla MGP'lere ait betimsel istatistikler hesaplanmış, ardından da normallik testi yapılmış olup sonuçlar aşağıda Tablo 18'de özetlenmiştir.

Tablo 18: MGP'lere Ait Betimsel İstatistikler ve Normallik Testi Sonuçları

Betimsel İstatistikler	Angoff Yöntemi	Yes/No Yöntemi	Ebel Yöntemi
N	17	17	17
Ortalama	51,47	57,94	47,12
Ortanca	53,25	55	48,5
Tepedeğer	24,75 ^a	55	58,25
Standart Sapma	12,41	15,42	15,19
Varyans	153,96	237,68	230,60
Çarpıklık Katsayısı	-0,85	0,49	-0,62
Basıklık Katsayısı	-0,09	-0,41	-0,46
Shaipro Wilk İstatistiği	0,91	0,96	0,92
Shaipro Wilk Sig.	0,10	0,50	0,15

a: Çoklu mod mevcut ve en küçük olan gösteriliyor.

Tablo 18'e göre yöntemlerin ortalama, ortanca ve mod değerleri birbirine yakındır. Angoff ve Yes/No yöntemlerinde çoklu modlar yer almaktadır. Çarpıklık katsayıları ise 0'dan çok uzakta değerler değildir. Basıklık katsayılarına bakıldığında da bu değerlerin -1 ve +1 arasında değerler olduğu görülmektedir. Bu bulgulara göre üç yöntem için de elde edilmiş olan kesme puanları dağılımının normal dağılıma yakın olduğu söylenebilir. Shapiro Wilk testi, veri sayısının 50'den az olduğu durumlarda kullanılan bir normallik testidir (Büyüköztürk, 2009, s. 42). Bu sınamaya göre de yöntemlere ait MGP'ler normal dağılımdan anlamlı bir fark göstermemektedir.

Betimsel istatistikler ve normallik testi sonuçları ışığında MGP'lerin normal dağılım gösterdiği söylenebilir. Bu nedenle analizde parametrik yöntemlere yer verilmiştir.

MGP'ler bakımından yöntemler arasındaki uyum için parametrik bir yöntem olan Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmıştır. Tablo 19'da yöntemlere ait MGP'ler arasındaki ilişkiler ve anlamlılık düzeyleri gösterilmiştir.

Tablo 19: MGP'ler Arasındaki Pearson Korelasyon Katsayıları

	Angoff Yöntemi	Yes/No Yöntemi	Ebel Yöntemi
Angoff Yöntemi	-	0,50*	0,83**
Yes/No Yöntemi	-	-	0,47

*p < 0,05

**p < 0,01

Tablo 19'a göre Angoff ve Ebel yöntemlerine ait MGP'ler arasında 0,01 hata düzeyinde yüksek bir ilişkinin varlığı söz konusudur. Bu %99 olasılıkla anlamlı değerlerin yüksek olmasının nedeni iki yöntemde de sınır gruba ait doğru cevaplama yüzdelerinin kullanılıyor olması olabilir. Angoff yönteminde tüm maddeler için bu yüzdeler ile kararlar verilirken; Ebel yönteminde maddeler Zorluk ve Uygunluk düzeylerine göre sınıflandırıldıktan sonra düzeyler için yüzdeler verilmektedir. Ayrıca Ebel yönteminde düzeyler için verilen yüzdeler Angoff yöntemindeki gibi madde bazında ele alındığında ise iki yöntemle ait MGP'ler arasındaki ilişkinin 0,01 manidarlık düzeyinde 0,82 olarak hesaplandığı görülmüştür.

MGP'ler arasındaki diğer bir manidar ilişkinin 0,05 hata düzeyinde Angoff ve Yes/No yöntemleri arasında olduğu görülmektedir. Buna göre, Angoff yönteminin bir varyasyonu olan Yes/No yöntemi ile Angoff yöntemi arasında MGP'ler bakımından, Ebel ve Angoff yöntemleri arasındaki ilişkiden daha düşük, orta derecede bir ilişki olduğu yorumu yapılabilir.

Tablo 19'a göre Yes/No ve Ebel yöntemlerine ait MGP'ler arasındaki 0,47 değerindeki ilişki anlamlı değildir. Bu durumun nedeni Yes/No yönteminde uzmanların sadece bir sınır grup öğrencisi düşünerek maddeler hakkında karar vermeleri olabilir.

Hesaplanan Pearson Korelasyon Katsayısı ortalamalardan bağımsız bir hesaplama yaptığı için yöntemlere ait MGP'lerin ortalamaları bağımlı gruplar t testi ile karşılaştırılmıştır ve sonuçları Tablo 20'de sunulmuştur.

Tablo 20: MGP Ortalamaları Arasındaki Farkın Testi

Karşılaştırılan Yöntemler	\bar{X}	t
Angoff Yöntemi	51,47	-1,89
Yes/No Yöntemi	57,94	
Angoff Yöntemi	51,47	2,13*
Ebel Yöntemi	47,12	
Yes/No Yöntemi	57,94	2,82*
Ebel Yöntemi	47,12	

*p < 0,05

Tablo 20'ye göre Angoff ve Yes/No yöntemlerine ait MGP'lerin ortalamaları sırayla 51,47 ve 57,94'tür. Uygulanan t testi sonucuna göre bu iki ortalama arasındaki farkın manidar düzeyde bir farklılık olmadığı söylenebilir. Angoff ve Ebel yöntemlerine ait MGP'lerin ortalamaları sırasıyla 51,47 ve 47,12'dir. Analiz sonuçlarına göre bu iki ortalama arasındaki farkın 0,05 hata düzeyinde, yani %95 olasılıkla anlamlı olduğu görülmektedir. Yes/No ve Ebel yöntemlerine ait MGP'lerin ortalamaları sırasıyla 57,94 ve 47,12'dir. Bu iki ortalama arasındaki farkın da 0,05 hata düzeyinde, yani %95 olasılıkla anlamlı olduğu görülmektedir.

Analiz sonuçları Angoff ve Yes/No yöntemleri ile elde edilen MGP'ler arasında pozitif ve orta düzeyde bir ilişki olduğunu, iki yönteme ait MGP'lerin ortalamaları arasında ise manidar bir farklılığın olmadığını göstermektedir. Bu iki bulguya göre Angoff ve Yes/No yöntemleri arasında MGP'ler bakımından uyum olduğu söylenebilir.

Angoff ve Ebel yöntemleri ile elde edilen MGP'ler arasında pozitif ve yüksek düzeyde bir ilişki olduğu, iki yönteme ait MGP'lerin ortalamaları arasında ise manidar bir farklılığın olduğu görülmektedir. Bu iki bulguya göre Angoff ve Ebel yöntemleri

arasında MGP'ler bakımından uyum olduğu ancak MGP'lerin ortalamaları bakımından farklılıkların söz konusu olduğu söylenebilir.

Yes/No ve Ebel yöntemleri ile elde edilen MGP'ler arasında ise anlamlı bir ilişkinin olmadığı ve iki yönteme ait MGP'lerin ortalamaları arasında ise manidar bir farklılığın olduğu görülmektedir. Bu iki bulguya göre Yes/No ve Ebel yöntemleri arasında MGP'ler bakımından uyum olmadığı söylenebilir.

3.4. Alt Problem 4: Yöntemlere ait uzman kararları arasındaki uyum nasıldır?

Bu alt problemin çözümünde uzman kararlarının yapısı dikkate alınarak Angoff ve Ebel yöntemleri için Kendall'ın W istatistiğinden ve Yes/No yöntemi için Cohen'in Kappa istatistiğinden faydalanılmıştır. Analiz sonuçlarına göre Angoff yönteminde de Ebel yönteminde de uzmanlar arasında 0,01 hata düzeyinde manidar bir uyum olduğu görülmüştür. Angoff yönteminde uzmanların uyumu için belirlenen W katsayısı 0,38 iken, Ebel yönteminde kararların maddeler bazında ele alınmasıyla elde edilen W değerinin 0,32 olduğu görülmüştür. Bu durumda uzmanlar arasında iki yöntemde de yüksek olmasa da kabul edilebilir bir uyum olduğu söylenebilir. Ayrıca Ebel yönteminde 3×4 'lük düzeyler için Kendall W 'si 0,01 hata düzeyinde 0,46 olarak hesaplanmıştır. Bu değer Angoff ve Ebel yöntemlerindeki uzmanlar arası uyumdan daha yüksek olduğu görülmektedir. Yes/No yöntemine ait uzman kararları için Cohen'in Kappa istatistiği kullanılmış ve tüm kararların uyuşma yüzdesi 0,60 olarak hesaplanmıştır. Bu değer de orta derecenin üstünde makul bir uyum olarak yorumlanabilir.

3.5. Alt Problem 5: Yöntemlerin varyans bileşenleri nasıldır?

Bu bölümde öncelikle yöntemlere göre elde edilen uzman kararlarının varyans bileşenleri her bir yöntem için ayrı ayrı hesaplanmıştır. Ardından tüm yöntemlerin birlikte ele alınması durumunda uzman kararlarına ait varyans bileşenleri belirlenmiştir.

3.5.1. Angoff yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?

Angoff yöntemiyle elde edilen uzman kararlarının varyanslarını ve varyans yüzdelerini G çalışmasıyla belirlemek amacıyla, 20 maddeye (m) 17 uzman (u) tarafından verilen kararlara tek değişkenlik kaynaklı $m \times u$ çaprazlanmış model uygulanmıştır. EduG (2010) programı yardımıyla “m” ve “u” ana etkileri ile “m x u” ortak etkisine ait kestirilen varyans ve toplam varyansı açıklama yüzdeleri aşağıda Tablo 21’de yer almaktadır.

Tablo 21: Angoff yöntemine ait kestirilen varyans ve toplam varyansı açıklama yüzdeleri

Değişkenlik Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans	Varyans Yüzdesi
Madde (m)	45261,06	19	2382,16	123,09	22,3
Uzman (u)	49266,61	16	3079,16	139,48	25,3
m x u, e	88055,04	304	289,65	289,65	52,5
Toplam	182582,71	339			%100

Tablo 21’e göre madde ana etkisine ait kestirilen varyansın (123,09) toplam varyansın %22,3’ünü açıkladığı görülmektedir. Maddeler için kestirilen bu değer tüm varyans değerleri içinde en küçük olandır. Araştırmalarda maddelere ait varyans yüzdesinin büyük olması beklenir/istenir. Maddelere ait varyans yüzdesinin düşük çıkması, maddelere ait farklılıkların ortaya çıkarılmadığının göstergesi olabilir. Ayrıca bu durum için maddelerin benzer özellikleri ölçtüğü yorumu da yapılabilir.

Uzman ana etkisine ait kestirilen varyansın (139,48) toplam varyansın %25,3’ünü açıkladığı görülmektedir. Uzmanlar için kestirilen bu değer en büyük ikinci varyanstır. Uzman ana etkisi için varyans bileşenin sıfıra yakın olması uzmanların birbirine benzer puanlama yaptığını gösterir (Shavelson ve Webb, 1991). Araştırmalarda uzmanlara ait varyans yüzdesinin küçük olması beklenir/istenir. Uzmanlar için elde edilen varyans yüzdesinin büyük çıkmasına dayanarak, uzmanların *aşırı olmasa da* birbirlerinden farklı puanlamalar yaptıkları, bir başka deyişle uzmanlara ait puanlamalar arasındaki tutarlılığın düşük olduğu söylenebilir. Bu sonucun nedeni uzmanların bir araya gelmeden bireysel olarak karar vermeleri olabilir. Nitekim Norcini ve diğerlerinin

(1987) çalışmasında Angoff yöntemi ile buluşmadan verilen kararların tutarlılığının buluşma sırasında ve buluşma sonrasında verilen kararların tutarlılığından daha düşük olduğu sonucuna ulaşılmıştır.

Madde-uzman ortak etkisine ait kestirilen varyansın (289,65) toplam varyansın %52,5'ini açıkladığı görülmektedir. Bu durum hem uzmanlar tarafından maddelere ait güçlük düzeylerinin farklı algılandığı hem de hata varyansının büyük olduğu anlamına gelebilir.

3.5.2. Yes/No yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?

Yes/No yöntemiyle elde edilen uzman kararlarının varyanslarını ve varyans yüzdelerini G çalışmasıyla belirlemek amacıyla, 20 maddeye (m) 17 uzman (u) tarafından verilen kararlara tek değişkenlik kaynaklı $m \times u$ çaprazlanmış model uygulanmıştır. EduG (2010) programı yardımıyla “m” ve “u” ana etkileri ile “m x u” ortak etkisine ait kestirilen varyans ve toplam varyansı açıklama yüzdeleri aşağıda Tablo 22’de yer almaktadır.

Tablo 22: Yes/No yöntemine ait kestirilen varyans ve toplam varyansı açıklama yüzdeleri

Değişkenlik Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans	Varyans Yüzdesi
Madde (m)	18,74	19	0,98622	0,05	19,0
Uzman (u)	7,61	16	0,47537	0,01	5,8
m x u, e	56,51	304	0,18589	0,19	75,1
Toplam	82,86	339			% 100

Tablo 22’ye göre madde ana etkisine ait kestirilen varyansın (0,05) toplam varyansın %19’unu açıkladığı görülmektedir. Maddeler için kestirilen bu değer tüm varyans değerleri içinde en küçük ikinci varyanstır. Araştırmalarda maddelere ait varyans yüzdesinin büyük olması beklenir/istenir. Maddeler ait varyans yüzdesinin düşük çıkması maddelere ait farklılıkların ortaya çıkarılmadığının göstergesi olabilir. Ayrıca bu durum için maddelerin benzer özellikleri ölçtüğü yorumu da yapılabilir.

Uzman ana etkisine ait kestirilen varyansın (0,01) toplam varyansın % 5,8'ini açıkladığı görülmektedir. Uzmanlar için kestirilen bu değer en küçük varyanstır. Uzman ana etkisi için varyans bileşenin sıfıra yakın olması uzmanların birbirine benzer puanlama yaptığını gösterir (Shavelson ve Webb, 1991). Araştırmalarda uzmanlara ait varyans yüzdesinin küçük olması beklenir/istenir. Uzmanlar için elde edilen varyans yüzdesinin düşük çıkmasına dayanarak uzmanların birbirlerinden farklı puanlamalar yapmadıkları, bir başka deyişle uzmanlara ait puanlamalar arasında bir tutarlılığın olduğu söylenebilir.

Madde-uzman ortak etkisine ait kestirilen varyansın (0,19) toplam varyansın %75,1'ini açıkladığı görülmektedir. Madde-uzman ortak etkisi için kestirilen bu değer en büyük varyanstır. Bu durum hem uzmanlar tarafından maddelere ait güçlük düzeylerinin farklı algılandığı hem de hata varyansının büyük olduğu anlamına gelebilir.

3.5.3. Ebel yöntemi ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?

Ebel yöntemiyle elde edilen uzman kararlarının varyanslarını ve varyans yüzdesini G çalışmasıyla belirlemek amacıyla, 20 maddeye (m) 17 uzman (u) tarafından verilen kararlara tek değişkenlik kaynaklı $m \times u$ çaprazlanmış model uygulanmıştır. EduG (2010) programı yardımıyla “m” ve “u” ana etkileri ile “m x u” ortak etkisine ait kestirilen varyans ve toplam varyansı açıklama yüzdesi aşağıda Tablo 23’de yer almaktadır.

Tablo 23: Ebel yöntemine ait kestirilen varyans ve toplam varyansı açıklama yüzdesi

Değişkenlik Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans	Varyans Yüzdesi
Madde (m)	31426,18	19	1654,00929	85,08	17,0
Uzman (u)	69607,79	16	4350,48713	207,14	41,4
m x u, e	63136,32	304	207,68527	207,68	41,5
Toplam	164170,29	339			% 100

Tablo 23'e göre madde ana etkisine ait kestirilen varyansın (85,08) toplam varyansın %17'sini açıkladığı görülmektedir. Maddeler için kestirilen bu değer tüm varyans değerleri içinde en küçük olandır. Araştırmalarda maddelere ait varyans yüzdesinin büyük olması beklenir/istenir. Maddeler ait varyans yüzdesinin düşük çıkması maddelere ait farklılıkların ortaya çıkarılmadığının göstergesi olabilir. Ayrıca bu durum için maddelerin benzer özellikleri ölçtüğü yorumu da yapılabilir.

Uzman ana etkisine ait kestirilen varyansın (207,14) toplam varyansın %41,4'ünü açıkladığı görülmektedir. Uzmanlar için kestirilen bu değer en büyük ikinci varyanstır. Uzman ana etkisi için varyans bileşenin sıfıra yakın olması uzmanların birbirine benzer puanlama yaptığını gösterir (Shavelson ve Webb, 1991). Araştırmalarda uzmanlara ait varyans yüzdesinin küçük olması beklenir/istenir. Uzmanlar için elde edilen varyans yüzdesinin yüksek çıkmasına dayanarak, uzmanların birbirlerinden farklı puanlamalar yaptıkları, bir başka deyişle uzmanlara ait puanlamalar arasındaki tutarlılığın düşük olduğu söylenebilir.

Madde-uzman ortak etkisine ait kestirilen varyansın (207,68) toplam varyansın %41,5'ini açıkladığı görülmektedir. Madde-uzman ortak etkisi için kestirilen bu değer en büyük varyanstır. Bu durum hem uzmanlar tarafından maddelere ait güçlük düzeylerinin farklı algılandığı hem de hata varyansının büyük olduğu anlamına gelebilir.

3.5.4. Angoff, Yes/No ve Ebel yöntemleri ile elde edilen uzman kararlarının varyans bileşenleri nasıldır?

Üç yöntemle elde edilen uzman kararlarının varyanslarını ve varyans yüzdelerini G çalışmasıyla belirlemek amacıyla, 20 maddeye (m) 3 farklı yöntemle göre (y) 17 uzman (u) tarafından verilen kararlara $m \times y \times u$ tümüyle çaprazlanmış model uygulanmıştır. EduG (2010) programı yardımıyla "m", "y" ve "u" ana etkileri ile "m x y", "m x u", "y x u", "m x y x u" ortak etkilerine ait kestirilen varyans ve toplam varyansı açıklama yüzdeleri aşağıda Tablo 24'te yer almaktadır.

Tablo 24: Üç yöntemle ait kestirilen varyans ve toplam varyansı açıklama yüzdeleri

Değişkenlik Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Varyans	Varyans Yüzdesi
Madde (m)	50887,15	19	2678,27095	47,15	5,0
Yöntem (y)	543675,96	2	271837,97941	529,04	55,6
Uzman (u)	72144,30	16	4509,01863	70,59	7,4
m x y	25818,83	38	679,44278	33,38	3,5
m x u	83195,04	304	273,66788	91,22	9,6
y x u	46737,71	32	1460,55337	67,43	7,1
m x y x u, e	68052,84	608	111,92902	111,93	11,8
Toplam	890511,83	1019			% 100

Tablo 24'e göre madde ana etkisine ait kestirilen varyansın (47,15) toplam varyansın %5'ini açıkladığı görülmektedir. Maddeler için kestirilen bu değer tüm varyans değerleri içinde en küçük ikinci ve ana etkiler içinde de en küçük varyanstır. Araştırmalarda maddelere ait varyans yüzdesinin büyük olması beklenir/istenir. Bu durum maddelere ait farklılıkların ortaya çıkarılmadığının göstergesi olabilir. Ayrıca bu durum için maddelerin benzer özellikleri ölçtüğü yorumu da yapılabilir.

Yöntem ana etkisine ait kestirilen varyansın (529,04) toplam varyansın %55,6'sını açıkladığı görülmektedir. Yöntemler için kestirilen bu değer tüm varyans değerleri içinde en büyük varyanstır. Bu durum yöntemler arasında kesme puanı belirlemede büyük farklılıkların olduğunu göstermektedir.

Uzman ana etkisine ait kestirilen varyansın (70,59) toplam varyansın %7,4'ünü açıkladığı görülmektedir. Uzmanlar için kestirilen bu değer tüm varyans değerleri en küçük dördüncü ve ana etkiler içinde de en küçük ikinci varyanstır. Uzman ana etkisi için varyans bileşenin sıfıra yakın olması uzmanların birbirine benzer puanlama yaptığını gösterir (Shavelson ve Webb, 1991). Araştırmalarda uzmanlara ait varyans yüzdesinin küçük olması beklenir/istenir. Uzmanlar için elde edilen varyans yüzdesinin düşük çıkmasına dayanarak, uzmanların birbirlerinden farklı puanlamalar yapmadıkları, bir başka deyişle uzmanlara ait puanlamalar arasında bir tutarlılığın olduğu söylenebilir.

Madde-yöntem ortak etkisine ait kestirilen varyansın (33,38) toplam varyansın %3,5'ini açıkladığı görülmektedir. Madde-yöntem ortak etkisi için kestirilen bu değer en küçük varyanstır. Bu durum maddelerin bir yöntemden diğer yönetime çok farklı puanlanmadığı anlamına gelebilir.

Madde-uzman ortak etkisine ait kestirilen varyansın (91,22) toplam varyansın %9,6'sını açıkladığı görülmektedir. Madde-uzman ortak etkisi için kestirilen bu değer en büyük üçüncü ve ortak etkiler içinde de en büyük ikinci varyanstır. Bu durum belirli bir uzmanın tüm maddeler için aynı puanlamayı yapmadığı ya da belirli bir maddenin puanlanmasıyla diğer maddelerin puanlaması bakımından bağımlılık olduğunu gösterebilir.

Yöntem-uzman ortak etkisine ait kestirilen varyansın (67,43) toplam varyansın %7,1'ini açıkladığı görülmektedir. Yöntem-uzman ortak etkisi için kestirilen bu değer en büyük beşinci ve ortak etkiler içinde de en büyük üçüncü varyanstır. Bu durum uzmanların bir yöntemden diğer yönetime kararlı puanlama yapmadıkları anlamına gelebilir. Bu durum ayrıca uzmanların yöntemlere göre farklı puanlama yaptıkları şeklinde de yorumlanabilir.

Madde-yöntem-uzman ortak (artık) etkisine ait kestirilen varyansın (111,93) toplam varyansın %11,8'ini açıkladığı görülmektedir. Madde-yöntem-uzman ortak etkisi için kestirilen bu değer ortak etkiler içinde en büyük varyanstır. Tüm bileşenler içinde ise en büyük ikinci varyans değeridir. Bu değerinin sıfıra yakın olması, madde-yöntem-uzman birleşiminden kaynaklanan varyansın küçük olduğu ya da ölçülemeyen hata kaynaklarının az olduğu; sıfırdan büyük olması ise ölçülemeyen hata kaynaklarının fazla olduğu şeklinde yorumlanabilir (Shavelson ve Webb, 1991). Elde edilen bulgu, ölçülemeyen hata kaynaklarının büyük olduğunu gösterebilir.

3.6. Alt Problem 6: D çalışmasına göre yöntemlere ait en uygun puanlayıcı sayısı nedir?

Aşağıda D çalışması yardımıyla her bir yöntem için uzman sayısının 5'er azalması ve 5'er artması senaryolarına göre elde edilen G ve Phi katsayıları tek tek açıklanmış, yöntemlere ait en uygun puanlayıcı sayıları belirlenmiştir.

Brennan (2004), G ve Phi katsayılarının yeterli ölçütlerinin isteğe bağlı olarak değiştiğini ancak bazı araştırmacıların G ve Phi katsayılarının 0,80'den büyük olması durumunda "yüksek" olarak değerlendirilebileceğini ifade etmektedir (Aktaran: Atılğan, 2004). Bunun yanında Shavelson ve Webb (1991) ise G ve Phi katsayılarının en az 0,80 olması gerektiğini belirtmişlerdir. Bulgulara ait değerlendirmeler de bu ölçüt ışığında ele alınmıştır.

3.6.1. Angoff yönteminde en uygun puanlayıcı sayısı nedir?

Angoff yöntemi ile elde edilen uzman kararlarına ait G ve Phi katsayılarının, uzman sayısının 5'er azaltılıp 5'er artırılması ile elde edilen farklı sayıda uzman sayısına sahip senaryolarda nasıl değiştiği D çalışması ile EduG (2010) programından faydalanılarak hesaplanmış olup sonuçları aşağıda Tablo 25'te gösterilmiştir.

Tablo 25: Angoff yönteminde farklı uzman sayılarına ait G ve Phi Katsayıları

Puanlayıcı Sayısı											
7		12		17*		22		27		32	
G	Phi	G	Phi	G	Phi	G	Phi	G	Phi	G	Phi
0,75	0,67	0,84	0,77	0,88	0,83	0,90	0,86	0,92	0,89	0,93	0,90

*Araştırmada kullanılan puanlayıcı (uzman) sayısı

Tablo 25'e göre Angoff yönteminde 20 madde ve 17 uzmana ait G ve Phi katsayıları sırasıyla 0,88 ve 0,83 olarak kestirilmiştir. Uzman sayısının 10 azalması (7 olması) durumunda G ve Phi katsayıları sırasıyla 0,75 ve 0,67 olmaktadır. Bu iki değer 0,80'den küçük olmaları nedeniyle 7 uzman sayısının Angoff yönteminde kesme puanı belirlemede yeterli olmadığı sonucuna varılabilir. Uzman sayısının 12 olması

durumunda ise G ve Phi katsayılarının sırasıyla 0,84 ve 0,77 olarak hesaplandığı görülmektedir. Bu iki değere göre mutlak değerlendirmelerde göz önünde bulundurulması gereken Phi katsayısının 0,80'den küçük olması nedeniyle Angoff yöntemiyle kesme puanı belirlenirken 12 uzmanın yeterli olmadığı yorumu yapılabilir.

G ve Phi katsayıları uzman sayısının 5 artırılması durumunda (22 uzman) 0,90 ve 0,86; 10 artırılması durumunda (27 uzman) 0,92 ve 0,89; 15 artırılması durumunda ise (32 uzman) 0,93 ve 0,90 olarak hesaplanmaktadır. Bu artışlar arasındaki farklılıkların çok büyük olmadığı görülmektedir. Bu durumda Angoff yöntemi ile kesme puanı belirlenirken 17 civarında uzmandan karar almanın yeterli olacağı, uzman sayısının artırılması ile güvenilirliğin fazla değişmeyeceği yorumu yapılabilir. Ayrıca bu durumun zaman ve emek açısından da ekonomik olmayacağı açıktır. Hurtz ve Hertz (1999) çalışmalarında Angoff yöntemi ile kesme puanı belirlenirken uygun uzman sayısının 10-15 civarında olması gerektiğini ifade etmişlerdir.

3.6.2. Yes/No yönteminde en uygun puanlayıcı sayısı nedir?

Yes/No yöntemi ile elde edilen uzman kararlarına ait G ve Phi katsayılarının, uzman sayısının 5'er azaltılıp 5'er artırılması ile elde edilen farklı sayıda uzman sayısına sahip senaryolarda nasıl değiştiği D çalışması ile EduG (2010) programından faydalanılarak hesaplanmış olup sonuçları Tablo 26'da gösterilmiştir.

Tablo 26: Yes/No yönteminde farklı uzman sayılarına ait G ve Phi Katsayıları

Puanlayıcı Sayısı											
7		12		17*		22		27		32	
G	Phi	G	Phi	G	Phi	G	Phi	G	Phi	G	Phi
0,64	0,62	0,75	0,74	0,81	0,80	0,85	0,84	0,87	0,86	0,89	0,88

*Araştırmada kullanılan puanlayıcı (uzman) sayısı

Tablo 26'ya göre Yes/No yönteminde 20 madde ve 17 uzmana ait G ve Phi katsayıları sırasıyla 0,81 ve 0,80 olarak kestirilmiştir. Uzman sayısının 10 azalması (7 olması) durumunda G ve Phi katsayıları sırasıyla 0,64 ve 0,62 olmaktadır. Bu iki değer 0,80'den küçük olmaları nedeniyle 7 uzman sayısının Angoff yönteminde kesme puanı

belirlemede yeterli olmadığı sonucuna varılabilir. Uzman sayısının 12 olması durumunda ise G ve Phi katsayılarının sırasıyla 0,75 ve 0,74 olarak hesaplandığı görülmektedir. Bu iki değerin 0,80'den küçük olmaları nedeniyle 12 uzman sayısının da Angoff yönteminde kesme puanı belirlemede yeterli olmadığı sonucuna varılabilir.

G ve Phi katsayıları uzman sayısının 5 artırılması durumunda (22 uzman) 0,85 ve 0,84; 10 artırılması durumunda (27 uzman) 0,87 ve 0,86; 15 artırılması durumunda ise (32 uzman) 0,89 ve 0,88 olarak hesaplanmaktadır. Bu artışlar arasındaki farklılıkların çok büyük olmadığı görülmektedir. Bu durumda Yes/No yöntemi ile kesme puanı belirlenirken 17 civarında uzmandan karar alınmasının yeterli olacağı, uzman sayısının artırılması ile güvenirliliğin fazla değişmeyeceği yorumu yapılabilir. Ayrıca bu durumun zaman ve emek açısından da ekonomik olmayacağı açıktır.

3.6.3. Ebel yönteminde en uygun puanlayıcı sayısı nedir?

Ebel yöntemi ile elde edilen uzman kararlarına ait G ve Phi katsayılarının, uzman sayısının 5'er azaltılıp 5'er artırılması ile elde edilen farklı sayıda uzman sayısına sahip senaryolarda nasıl değiştiği D çalışması ile EduG (2010) programından faydalanılarak hesaplanmış olup sonuçları Tablo 27'de gösterilmiştir.

Tablo 27: Ebel yönteminde farklı uzman sayılarına ait G ve Phi Katsayıları

Puanlayıcı Sayısı											
7		12		17*		22		27		32	
G	Phi	G	Phi	G	Phi	G	Phi	G	Phi	G	Phi
0,74	0,59	0,83	0,71	0,87	0,78	0,90	0,82	0,92	0,85	0,93	0,87

*Araştırmada kullanılan puanlayıcı (uzman) sayısı

Tablo 27'ye göre Ebel yönteminde 20 madde ve 17 uzmana ait G ve Phi katsayıları sırasıyla 0,87 ve 0,78 olarak kestirilmiştir. Bu iki değere göre mutlak değerlendirmelerde göz önünde bulundurulması gereken Phi katsayısının 0,80'den küçük olması nedeniyle Ebel yöntemiyle kesme puanı belirlenirken 17 uzmanın yeterli olmadığı yorumu yapılabilir.

Uzman sayısının 10 azalması (7 olması) durumunda G ve Phi katsayıları sırasıyla 0,74 ve 0,59 olmaktadır. Bu iki değerin 0,80'den küçük olmaları nedeniyle 7 uzman sayısının Ebel yönteminde kesme puanı belirlemede yeterli olmadığı sonucuna varılabilir. Uzman sayısının 12 olması durumunda ise G ve Phi katsayılarının sırasıyla 0,83 ve 0,71 olarak hesaplandığı görülmektedir. Bu iki değere göre mutlak değerlendirmelerde göz önünde bulundurulması gereken Phi katsayısının 0,80'den küçük olması nedeniyle Ebel yöntemiyle kesme puanı belirlenirken 12 uzmanın yeterli olmadığı yorumu yapılabilir. G ve Phi katsayıları uzman sayısının 5 artırılması durumunda (22 uzman) 0,90 ve 0,82; 10 artırılması durumunda (27 uzman) 0,92 ve 0,85; 15 artırılması durumunda ise (32 uzman) 0,93 ve 0,87 olarak hesaplanmaktadır. Bu durumda 22 uzman sayısının Ebel yöntemi ile kesme puanı belirlemede yeterli olduğu sonucuna ulaşılmaktadır. Ayrıca 27 ve 32 uzman sayısındaki G ve Phi katsayıları arasındaki artışların çok büyük olmadığı görülmektedir. Bu durumda Ebel yöntemi ile kesme puanı belirlenirken 22 civarında uzmandan karar almanın yeterli olacağı, uzman sayısının artırılması ile güvenilirliğin fazla değişmeyeceği yorumu yapılabilir. Ayrıca bu durumun zaman ve emek açısından da ekonomik olmayacağı açıktır.

BÖLÜM IV

SONUÇ VE ÖNERİLER

Bu bölümde her bir alt problem için ulaşılan sonuçlar ve bu sonuçlara dayalı olarak getirilebilecek öneriler yer alacaktır.

4.1. SONUÇLAR

Araştırmada elde edilen sonuçlar alt problemlerin sırasına uygun olarak aşağıda özetlenmiştir.

1. Angoff, Yes/No ve Ebel yöntemleri ile belirlenen kesme puanlarının birbirinden farklı olduğu görülmektedir. En yüksek kesme puanı Yes/No yöntemi ile elde edilirken, Yes/No yöntemini Angoff yöntemi takip etmektedir. En düşük kesme puanı ise Ebel yöntemi ile belirlenmiştir. Kesme puanlarındaki bu farklılık başarılı sayılan öğrenci sayısını ve yüzdesini etkilemiştir. Bu üç yöntemle belirlenen kesme puanlarına göre başarılı sayılan öğrenci yüzdeleri arasında manidar bir farklılık olduğu sonucuna ulaşılmıştır.
2. Öğrenci puanlarının hesaplanan üç kesme puanı dikkate alınarak başarılı-başarısız durumlarına göre “1-0” kodlanması sonucunda yöntemler arasında manidar düzeyde uyum olduğu sonucuna varılmıştır. Angoff – Yes/No ve Angoff – Ebel yöntemleri arasında mükemmel derecede bir uyum görülürken, Yes/No – Ebel yöntemleri arasında makul bir uyumun olduğu sonucuna ulaşılmıştır.
3. Elde edilen MGP’ler bakımından Angoff ve Ebel yöntemleri arasında yüksek derecede bir uyumun olduğu ancak MGP ortalamalarının manidar düzeyde farklı olduğu; Angoff ve Yes/No yöntemleri arasında ise orta derecede bir uyumun olduğu ve MGP ortalamalarının manidar düzeyde farklılaşmadığı görülürken; Yes/No yöntemi ve Ebel yöntemi arasında anlamlı bir uyumun olmadığı ve MGP ortalamalarının manidar düzeyde farklılaştığı sonucuna ulaşılmıştır.

4. Yes/No yönteminde uzmanlar arası uyumun orta derecenin üstünde makul bir uyum olduğu; Angoff ve Ebel yöntemlerinde ise uzmanlar arasında manidar bir uyum olduğu ve Ebel yöntemindeki kararların 3x4'lük düzeyler ile hesaplanması durumunda ise uzmanlar arası uyumun Angoff yönteminden daha yüksek olduğu sonucuna ulaşılmıştır.
5. G çalışması sonucunda uzman kararlarının varyans bileşenlerine göre
 - a. Üç yöntemin maddelere ait farklılıkları fazla ortaya çıkaramadığı ya da maddelerin benzer özellikleri ölçtüğü; Ebel yönteminin uzmanlar arası tutarlılık bakımından en düşük sonucu verdiği, onu takiben Angoff yönteminin geldiği ve en yüksek uzman tutarlılığını Yes/No yönteminin verdiği sonucuna ulaşılmıştır. Ayrıca üç yöntemde de hem uzmanlar tarafından maddelere ait güçlük düzeylerinin farklı algılandığı hem de hata varyansının büyük olduğu görülmüştür.
 - b. Üç yöntemin birlikte ele alındığı G çalışması sonucunda maddeler arası farklılığın ortaya çıkarılamadığı ya da maddelerin benzer özellikleri ölçtüğü; uzmanlar arasında bir uyumun olduğu ve yöntemler arasında da büyük bir farklılığın olduğu sonucuna ulaşılmıştır. Yöntemler arasındaki bu büyük farklılığa dayanarak kesme puanlarının yöntemden çok etkilendiğini, bu etkinin yöntemlerin kesme puanı belirlemedeki yaklaşım farklılıklarından kaynaklandığını söyleyebiliriz. Ayrıca maddelerin bir yöntemden diğer yönteme çok farklı puanlanmadığı; belirli bir uzmanın tüm maddeler için aynı puanlamayı yapmadığı ya da belirli bir maddenin puanlanmasıyla diğer maddelerin puanlaması bakımından bağımlılık olduğu; uzmanların bir yöntemden diğer yönteme kararlı puanlama yapmadıkları, bir başka deyişle uzmanların yöntemlere göre farklı puanlama yaptıkları ve ölçülemeyen hata kaynaklarının büyük olduğu sonucuna ulaşılmıştır.
6. Yapılan D çalışması sonucuna göre Angoff ve Yes/No yöntemleri ile kesme puanı belirlenirken 17 civarında uzmandan; Ebel yönteminde ise 22 civarında

uzmandan karar almanın güvenilirlik için yeterli olacağı sonucuna ulaşılmıştır. Daha fazla sayıdaki uzman ile G ve Phi katsayılarının arttığı ancak bu katsayılarda yine de büyük bir değişme olmadığı görülmüştür.

4.2. ÖNERİLER

4.2.1. Araştırma Sonuçlarından Çıkan Öneriler

- Bir çalışmada Angoff, Yes/No ve Ebel yöntemlerinden yalnızca biri kullanılacaksa G ve Phi katsayılarının daha yüksek olması nedeniyle Angoff yönteminin tercih edilmesi önerilebilir.
- Ebel yönteminin kullanılacağı çalışmalarda uzman sayısının 20'den fazla olmasına dikkat edilmesi önerilebilir. Angoff ve Yes/No yöntemleri kullanılacaksa bu sayının 17 civarında olması yeterli olacaktır.
- Angoff ve Yes/No yöntemlerinden elde edilen MGP'lerin birbiriyle uyumlu olması nedeniyle iki yöntemin beraber kullanılmasının zaman bakımından ekonomik olmayacağı durumlarda sadece biri seçilebilir.
- Yöntemler arasında belirli bir uyum olsa da başarılı sayılan öğrenci yüzdeleri bakımından manidar farklılıklar olduğu göz önünde bulundurulmalıdır.

4.2.2. Araştırmacılara Yönelik Öneriler

- Bu çalışmada Angoff, Yes/No ve Ebel standart belirleme yöntemleri kullanılmıştır. Benzer bir çalışma başka yöntemlerle yapılabilir.
- Ebel yönteminde elde edilen yüzdeler madde bazında ele alınıp Angoff yöntemindeki yüzdeler ile karşılaştırılabilir.
- Uzmanlardan bu üç yöntem için farklı zamanlarda karar alınabilir. Aynı anda alınan kararlarla diğerleri karşılaştırılabilir. Aralarındaki uyuma ve farka bakılabilir.
- Bu üç yönteme, öğrencilerin seçenekleri elemesiyle doğru cevabı bulduğu varsayımına dayanan Nedelksy yöntemi de dâhil edilebilir. Yöntemler arası karşılaştırmalar çoğaltılabilir.

- Bu yöntemler, sınır-grup yöntemi gibi grubun performansına dayanan standart belirleme yöntemleriyle karşılaştırılabilir.
- Uzmanların alanlarındaki deneyim durumlarına göre (görev yaptıkları yer, çalıştıkları yıl vb.) verdikleri kararlar karşılaştırılabilir.
- Uç değerlerin analiz dışına çıkarılması durumunda kesme puanlarının nasıl değiştiği, başarılı sayılan öğrenci oranını etkileyip etkilemediği ve G çalışması sonucu varyans bileşenlerine ait değerlerin nasıl olduğu incelenebilir.
- Yöntemlerin hangisinin daha kolay-uygulanabilir olduğu uzmanlardan görüşler alınarak nitel bir çalışmayla belirlenebilir.
- Belirlenen standartların ülkemiz okullarında 45 olan geçme notu ile ilişkisine bakılabilir.
- Yöntemlerin sınıflama tutarlığına bakılabilir.

KAYNAKÇA

- Andrew, B. J. ve Hecht, J. T. (1976). A Preliminary Investigation of Two Procedures for Setting Examination Standards. *Educational and Psychological Measurement*, 36, 45 – 50.
- Appleton, L. P. (1984). *A Comparison of the Nedelsky, Ebel, and Contrasting Groups Methods of Cut Score Determination*. Yayınlanmamış doktora tezi, The Catholic University of America, America.
- Atılğan, H. (2004). *Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma*. Yayınlanmamış Doktora tezi, Hacettepe Üniversitesi, Ankara.
- Atılğan, H., Kan, A., Doğan, N. (2009). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Anı Yayıncılık.
- Baykul, Y. (2000) *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM Yayınları.
- Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlag.
- Brennan, R. L. ve Lockwood, R. E. (1980). A Comparison of the Nedelsky and Angoff Cutting Score Procedures Using Generalizability Theory. *Applied Psychological Measurement*, 4, 219 – 240.
- Buckhendal, C. W., Smith, R. W., Impara, J. C. ve Plake, B. S. (2002). A Comparison of Angoff and Bookmark Standard Setting Methods. *Journal of Educational Measurement*, 39 (3), 253 – 263.
- Büyüköztürk, Ş. (2009). *Sosyal Bilimler İçin Veri Analizi El Kitabı*. Ankara: Pegem Akademi.

- Cardinet, J., Johnson, S., Pini, G. (2010). *Applying Generalizability using EduG*. New York: Taylor & Francis Group.
- Chang, L. (1996). Comparison Between The Nedelsky and Angoff Standart-Setting Methods. Paper presented at the Annual Meeting of the National Council on Measurement in Education. New York, USA.
- Chinn, R. N. ve Hertz, N. R. (2002), Alternative Approaches to Standard Setting for Licensing and Certification Examinations. *Applied Measurement in Education*, 15 (1), 1 – 14.
- Cizek, G. J. (2001). *Setting Performance Standards: Concepts, Methods and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cizek, G. J ve Bunch, M. B. (2007). *Standard Setting: A Guide to Establishing and Evaluating*. California: Sage Publications.
- Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Florida: Holt, Rinehart and Winston, Inc.
- Cusimano, M. D. ve Rothman, A. I. (2003). The Effect of Incorporating Normative Data into a Criterion-Referenced Standard Setting in Medical Education. *Academic Medicine*, 78 (10), 88 - 90
- Çetin, S. (2011). *İşaretleme ve Angoff Standart Belirleme Yöntemlerinin Karşılaştırılması*. Yayınlanmamış Doktora tezi, Hacettepe Üniversitesi, Ankara.
- Deliceoğlu, G. (2009). *Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Genellenebilirlik ve Klasik Test Kuramına Dayalı Güvenirliklerinin Karşılaştırılması*. Yayınlanmamış Doktora tezi, Ankara Üniversitesi, Ankara.
- Demirel, Ö. (2005). *Öğretme Sanatı*. Ankara: Pegem A Yayıncılık.

- Downing, S. M., Tekian, A. ve Yudkowsky, R. (2006). Research Methodology: Procedures for Establishing Defensible Absolute Passing Scores on Performance Examinations in Health Professions Education. *Teaching and Learning in Medicine*, 18 (1), 50 – 57.
- Ebel, R. L. (1972). *Essentials of Educational Measurement*. New Jersey: Prentice Hall.
- Fergusson, G. A. (1971). *Statistical Analysis in Psychology & Education*. USA: McGraw-Hill.
- Güler, N. (2008). *Klasik Test Kuramı Genellenabilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma*. Yayınlanmamış Doktora tezi, Hacettepe Üniversitesi, Ankara.
- Hambleton, R. K. (2001). Setting Performance Standards on Educational Assessments and Criteria for Evaluating The Process. G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods and Perspectives* (s. 89 – 116). Mahwah, NJ: Lawrence Erlbaum Associates.
- Harasym, P. H. (1981). A Comparison of the Nedelsky and Modified Angoff Standard Setting Procedure On Evaluation Outcome. *Educational and Psychological Measurement*, 41, 725 – 734.
- Howell, D. C. (2007). *Statistical Methods for Psychology*. USA: Thomson Learning Academic Resource Center.
- Hurtz G. M. ve Hertz N. R. (1999). How Many Raters Should be Used for Establishing Cutoff Scores with the Angoff Method? a Generalizability Theory Study. *Educational And Psychological Measurement*, 59, 885 – 897.
- Impara, J. C. ve Plake, B. S. (1997). Standard Setting: An Alternative Approach. *Journal of Educational Measurement*, 34 (4), 353 – 366.

- Impara, J. C. ve Plake, B. S. (1998). Teachers' Ability to Estimate Item Difficulty: A Test of the Assumptions in the Angoff Standard Setting Method. *Journal of Educational Measurement*, 35 (1), 69 – 81.
- Kaptan, S. (1977). *Bilimsel Araştırma Teknikleri*. Ankara: Tekişik Matbaası ve Rehber Yayınevi.
- Karasar, N. (1982). *Bilimsel Araştırma Yöntemi*. Ankara: Matbaş Matbaacılık ve Ambalaj Sanayi.
- Livingston, S. A. ve Zieky, M. J. (1982). Passing Scores (A Manual For Setting Standards Of Performance On Educational And Occupational Tests) [Elektronik Sürüm]. New Jersey: Educational Testing Service. Erişim: 12 Kasım 2009, <http://www.ets.org/portal/site/ets/menuitem.c988ba0e5dd572bada20bc47c3921509/?vgnextoid=0dcfd639ef4a9010VgnVCM10000022f95190RCRDvevgnnextchannel=dcb3be3a864f4010VgnVCM10000022f95190RCRD>
- Livingston, S. A. ve Zieky, M. J. (1983). *A Comparative Study of Standard-Setting Methods* (Rapor No: ETS-RR-83-38). Princeton: Educational Testing Service.
- Mills, C. N. (1983). A Comparison of Three Methods of Establishing Cut-off Scores on Criterion-Referenced Tests. *Journal of Educational Measurement*, 20 (3), 283 -292.
- Mills, C. N. ve Berr, J. E. (1983). A Comparison of Standard Setting Methods: Do The Same Judges Establish The Same Standards With Different Methods? Paper presented at the Annual Meeting of the American Educational Research Association: Montreal, Canada.
- Mills, C. N, ve Melican, G. J. (1988). Estimating and Adjusting Cutoff Scores: Features of Selected Methods. *Applied Measurement in Education*, 1, 261 – 275.

- Murphy, K. R.; Davidshofer, C. O. (1991). *Psychological Testing: Principles and Applications*. New Jersey: Prantice Hall.
- Norcini, J. J., Lipner, R. S., Langdon L. O. ve Strecker, C. A. (1987). A Comparison of Three Variations on a Standard-Setting Method. *Journal Of Educational Measurement, 24 (1)*, 56 – 64.
- Ömür, S. ve Selvi, H. (2010). Angoff, Ebel ve Nedelsky Yöntemleriyle Belirlenen Kesme Puanlarının Sınıflama Tutarlılıklarının Karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1 (2)*, 109 – 113.
- Poggio, J. P., Glasnapp D. R. ve Eros P. I. (1982). An Evaluation of Contrasting Groups Methods for Setting Standards. Paper presented at The Annual Meeting of the American Educotional Research Association. New York: USA.
- Shavelson, R. J. ve Webb, N. M. (1991). *Generalizability Theory: A Primer*. California: SAGE Publications.
- Skakun, E. N. ve Kling, S. (1980). Comparability of Methods for Setting Standards. *Journal of Educational Measurement, 17 (3)*, 229 – 235.
- Şencan, H. (2005). *Sosyal ve Davranışsal Ölçümlerde Güvenirlilik ve Geçerlilik*. Ankara: Seçkin Yayıncılık.
- Tanrıverdi, S. (2006). *Standart Belirleme Yöntemlerinin Geçme Puanları Üzerindeki Etkisi*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.
- Taşdelen, G. (2009). *Nedelsky ve Angoff Standart Belirleme Yöntemlerinin Genellenebilirlik Kuramı İle Karşılaştırılmasına İlişkin Bir Araştırma*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.
- Tekin, H. (1991). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Yargı Kitap ve Yayınevi.

Turgut, M. F. (1993). *Eđitimde Ölçme ve Deęerlendirme Metodları*. Ankara: Saydam Matbaacılık.

Zieky, M. J. (2001). So Much Has Changed: How the Setting of Cutscores Has Evolved Since the 1980s. G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods and Perspectives* (s. 19 – 51). Mahwah, NJ: Lawrence Erlbaum Associates.

EKLER

TESTLE İLGİLİ AÇIKLAMA

Bu test Matematik dersindeki öğrenme düzeyinizi belirlemek için hazırlanmıştır. Testte Matematik dersi “Bölme İşlemi ve Kesirler” konularıyla ilgili toplam 20 soru bulunmaktadır. Testin tümü için cevaplama süresi 35 dakikadır.

Her sorunun dört seçeneği vardır. Dört seçenekten yalnızca bir tanesi doğru cevaptır. Soruları dikkatlice okuyunuz ve doğru bulduğunuz seçeneği cevap kâğıtlarına işaretleyiniz.

Uyarı: Lütfen soruları boş bırakmamaya özen gösteriniz.

Başarılar diliyorum. 😊

Adı / Soyadı:

Sınıfı / Numarası:

SORULAR

1. $525 \div 21$ işleminin sonucu kaçtır?

a) 25

b) 52

c) 205

d) 250

2. Bahçeden toplanan 4913 kg elma 17 kilogramlık kasalara doldurulacaktır. Kaç tane kasa gerekir?

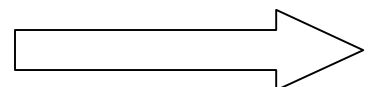
a) 208

b) 209

c) 288

d) 289

Diğer sayfaya geçiniz.



3. 60 elmanın $\frac{6}{10}$ 'u kaç elma eder?

a) 10

b) 36

c) 100

d) 600

4. $\frac{3}{5} + \frac{4}{5}$ işleminin sonucu kaçtır?

a) $1\frac{1}{5}$

b) $1\frac{2}{5}$

c) $1\frac{3}{5}$

d) $1\frac{4}{5}$

5. Aşağıdakilerden hangisi **yanlıştır**?

a) $\frac{3}{4} = \frac{6}{8}$

b) $\frac{8}{5} = 1\frac{3}{5}$

c) $\frac{7}{3} = 2\frac{1}{3}$

d) $\frac{5}{7} = \frac{6}{14}$

6. $(48 \div 12) \times (12 \div 4)$ işleminin sonucu kaçtır?

a) 12

b) 16

c) 18

d) 20

7. 861 lt. zeytinyağını 18 lt. lik tenekelere doldurduktan sonra kalan yağı 3lt. lik kaplara doldurmak istiyoruz. Toplam kaç tane teneke ve kap kullanmamız gerekir?

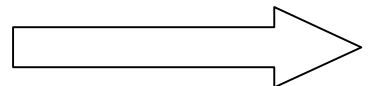
a) 5

b) 25

c) 47

d) 52

Diğer sayfaya geçiniz.



8. Aşağıdakilerden hangisinde “>” ve “<” sembolleri doğru olarak kullanılmıştır?

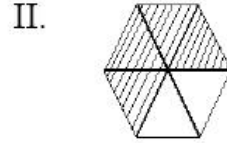
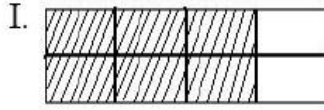
a) $3\frac{1}{8} < \frac{12}{8}$

b) $2\frac{3}{6} > \frac{20}{6}$

c) $1\frac{3}{5} < \frac{6}{5}$

d) $2\frac{3}{4} > \frac{10}{4}$

9.



Yukarıdaki şekillerden hangisi veya hangileri $\frac{3}{4}$ kesrini ifade eder?

a) Yalnız I

b) I ve III

c) Yalnız III

d) I, II ve III

10. $\frac{3}{12} + \frac{13}{12}$ işleminin sonucu kaçtır?

a) $\frac{16}{24}$

b) $1\frac{4}{12}$

c) $1\frac{8}{12}$

d) $\frac{10}{12}$

11. Ceyhan, sepetteki elmaların $\frac{3}{6}$ 'sını annesine, $\frac{2}{6}$ 'sını da arkadaşına veriyor. Geriye 24 tane elma kaldığına göre, başlangıçta sepette kaç tane elma vardı?

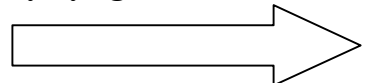
a) 32

b) 36

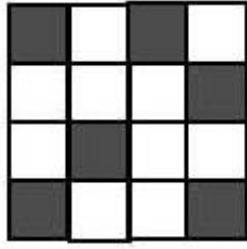
c) 44

d) 144

Diğer sayfaya geçiniz.



12. Aşağıdaki şekilde bir bütün eş parçalarına bölünmüştür. Verilen kesirlerden hangisi taralı kısımlara karşılık gelir?



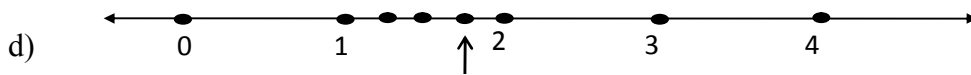
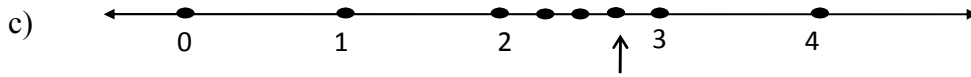
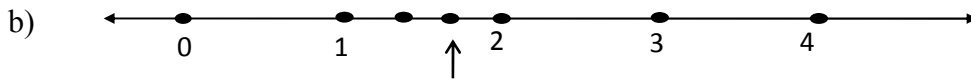
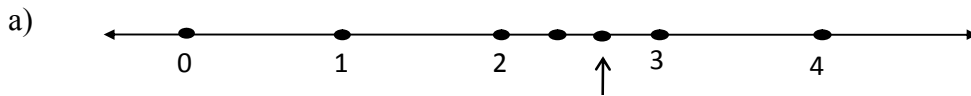
a) $\frac{4}{8}$

b) $\frac{4}{16}$

c) $\frac{5}{16}$

d) $\frac{6}{16}$

13. $1\frac{2}{3}$ kesrinin sayı doğrusu üzerindeki okla gösterimi aşağıdakilerden hangisidir?



14. $323 \div 17$ işleminin sonucu kaçtır?

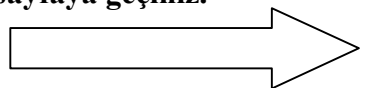
a) 15

b) 17

c) 18

d) 19

Diğer sayfaya geçiniz.



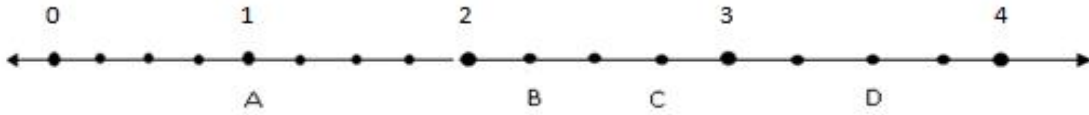
15. $\frac{A}{7} > \frac{8}{7}$ karşılaştırmasında “A” yerine yazılabilecek sayı hangisidir?

- a) 3 b) 6 c) 8 d) 9

16. Aşağıdaki sıralamalardan hangisi yanlıştır?

- a) $\frac{4}{5} < \frac{2}{5} < \frac{5}{5}$ b) $\frac{3}{2} < \frac{5}{2} < \frac{7}{2}$ c) $\frac{3}{5} < \frac{4}{5} < \frac{5}{5}$ d) $\frac{2}{3} < \frac{3}{3} < \frac{7}{3}$

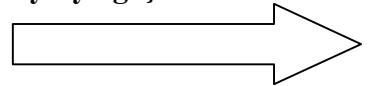
17.



Yukarıdaki sayı doğrusunda A, B, C, D harfleri ile gösterilen noktalar aşağıdakilerden hangisinde doğru olarak verilmiştir?

- a) $A = \frac{4}{4}$; $B = \frac{2}{4}$; $C = \frac{8}{4}$; $D = \frac{2}{4}$
b) $A = \frac{1}{4}$; $B = \frac{2}{4}$; $C = \frac{11}{4}$; $D = \frac{16}{4}$
c) $A = \frac{4}{4}$; $B = \frac{9}{4}$; $C = \frac{11}{4}$; $D = \frac{14}{4}$
d) $A = \frac{4}{4}$; $B = \frac{9}{4}$; $C = \frac{12}{4}$; $D = \frac{13}{4}$

Diğer sayfaya geçiniz.



18. $\frac{1}{7}$ ' si 28 olan sayı kaçtır?

a) 192

b) 194

c) 196

d) 198

19. $\frac{8}{9} - \frac{A}{9} = \frac{3}{9}$ işleminde "A" yerine aşağıdaki sayılardan hangisi yazılmalıdır?

a) 2

b) 5

c) 6

d) 7

20. Bir memur, maaşının $\frac{2}{6}$ ' sını kiraya; $\frac{1}{6}$ ' sını elektriğe ve $\frac{1}{6}$ ' sını da borçlarına vermiştir.

Bu memur maaşının kaçta kaçını harcamıştır?

a) $\frac{3}{6}$

b) $\frac{4}{6}$

c) $\frac{5}{6}$

d) $\frac{6}{6}$

Test bitti.

Yanıtlarınızı kontrol ediniz.

UZMAN GÖRÜŞLERİ İÇİN ARAŞTIRMA FORMU

Değerli öğretmenler,

Yüksek lisans tezi için yaptığım araştırmada, Matematik dersi – “Bölme ve Kesirler” alt öğrenme alanları kazanımlarına yönelik hazırlamış olduğum 20 soruluk çoktan seçmeli test için üç farklı değerlendirme yapmanıza ihtiyaç duyulmaktadır. Yaptığınız değerlendirmeler ile teste ait geçme puanı belirlenecektir. Araştırma kapsamında elde edilecek bilgilere göre gerçek hayatta öğrencilerin dersten kalması-geçmesi gibi bir durum söz konusu olmayacaktır. Bu nedenle değerlendirmelerinizi içtenlikle yapmanızı rica eder şimdiden hepinize gösterdiğiniz özveri için teşekkür ederim.

Hacettepe Üniversitesi Eğitim Bilimleri Bölümü

Arş. Gör. Ceylan Gündeğer

Değerlendirme 1: “Matematik dersinde geçme-kalma sınırında (minimum yeterlik düzeyinde) olan öğrencilerin yüzde kaç soruyu doğru cevaplayabilir?” sorusuna cevap vermeniz beklenmektedir.

Örnek: %80 gibi... (%80’i doğru cevaplayabilir ifadesi aynı zamanda sorunun kolay olduğuna işaret edecektir.)

Değerlendirme 2: “Matematik dersinde geçme-kalma sınırında (minimum yeterlik düzeyinde) olan öğrenci soruyu doğru cevaplayabilir mi?” sorusuna cevap vermeniz beklenmektedir. Verilenlerden hangisi size uygunsa altındaki kutucuğa çarpı işareti koyarak belirtiniz.

Örnek:

Doğru cevaplayabilir

Doğru cevaplayamaz

(X)

()

Değerlendirme 3: Bu değerlendirme türünde sizlerden iki karar vermeniz ve bu kararları aşağıdaki tabloya kendi fikirleriniz doğrultusunda yerleştirmeniz beklenmektedir.

Bu amaçla öncelikle tüm sorulara tekrar göz atmanızı tavsiye ederiz. Sorunun testteki diğer sorulara göre “**Zorluk**” derecesine karar veriniz: Soru sizce **kolay** mı, **orta** zorlukta mı yoksa **zor** mu? Ardından sorunun testteki diğer sorulara göre “**Uygunluk**” derecesine karar veriniz: Soru **gerekli** mi, **önemli** mi, **kabul edilebilir** (yani teste alınabilir) mi yoksa **tartışılabilir** mi (tartışmaya açık mı)? Bu iki karar göre sorunun numarasını uygun gördüğünüz hücrenin “*Sorular*” kısmına yazınız. Bütün hücreler dolmak zorunda değildir.

Son olarak da Matematik dersinde geçme-kalma sınırında (minimum yeterli düzeyinde) olan öğrencilerin yüzde kaç hücredeki soruları doğru cevaplayabilir?” sorusunu düşünerek “*Yüzde*” kısmına yazınız.


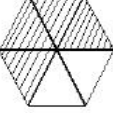
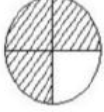
Örnek:

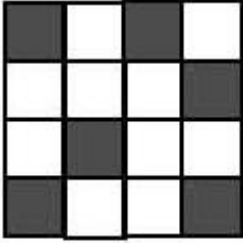
UYGUNLUK	ZORLUK		
	<i>Kolay</i>	<i>Orta</i>	<i>Zor</i>
<i>Gerekli</i>	<i>Sorular:1,5,8</i> <i>Yüzde:%60</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>
<i>Önemli</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>
<i>Kabul Edilebilir</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>
<i>Tartışılabilir</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>

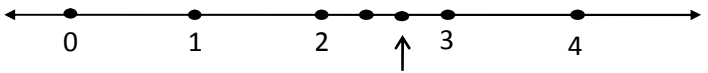
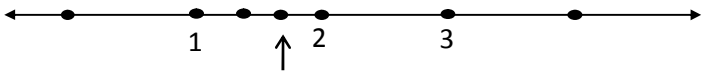
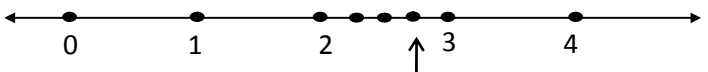
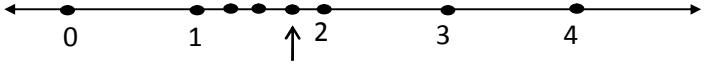
NOT: Değerlendirmelerinize test sorularına göz attıktan sonra başlayabilirsiniz. Değerlendirmelerde de, biri tamamen bitmeden diğerine geçmemenizi öneririz.

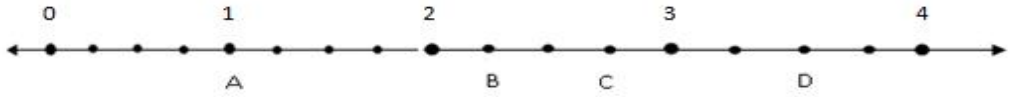
Kolay gelsin. 😊

DEĞERLENDİRME 1	SORULAR	DEĞERLENDİRME 2	
Yüzde kaç doğru cevaplayabilir? %	<p>5. Aşağıdakilerden hangisi yanlıştır?</p> <p>a) $\frac{3}{4} = \frac{6}{8}$ b) $\frac{8}{5} = 1 \frac{3}{5}$ c) $\frac{7}{3} = 2 \frac{1}{3}$ d) $\frac{5}{7} = \frac{6}{14}$</p>	Doğru cevaplayabilir ()	Doğru cevaplayamaz ()
Yüzde kaç doğru cevaplayabilir? %	<p>6. $(48 \div 12) \times (12 \div 4)$ işleminin sonucu kaçtır?</p> <p>a) 12 b) 16 c) 18 d) 20</p>	Doğru cevaplayabilir ()	Doğru cevaplayamaz ()
Yüzde kaç doğru cevaplayabilir? %	<p>7. 861 lt. zeytinyağını 18 lt. lik tenekelere doldurduktan sonra kalan yağı 3lt. lik kaplara doldurmak istiyoruz. Toplam kaç tane tenke ve kap kullanmamız gerekir?</p> <p>a) 5 b) 25 c) 47 d) 52</p>	Doğru cevaplayabilir ()	Doğru cevaplayamaz ()
Yüzde kaç doğru cevaplayabilir? %	<p>8. Aşağıdakilerden hangisinde “>” ve “<” sembolleri doğru olarak kullanılmıştır?</p> <p>a) $3 \frac{1}{8} < \frac{12}{8}$ b) $2 \frac{3}{6} > \frac{20}{6}$ c) $1 \frac{3}{5} < \frac{6}{5}$ d) $2 \frac{3}{4} > \frac{10}{4}$</p>	Doğru cevaplayabilir ()	Doğru cevaplayamaz ()

DEĞERLENDİRME 1	SORULAR	DEĞERLENDİRME 2	
<p>Yüzde kaçını doğru cevaplayabilir?</p> <p>%</p>	<p>9.</p> <p>I.  II.  III. </p> <p>Yukarıdaki şekillerden hangisi veya hangileri $\frac{3}{4}$ kesrini ifade eder?</p> <p>a) Yalnız I b) I ve III c) Yalnız III d) I, II ve III</p>	<p>Doğru cevaplayabilir</p> <p>()</p>	<p>Doğru cevaplayamaz</p> <p>()</p>
<p>Yüzde kaçını doğru cevaplayabilir?</p> <p>%</p>	<p>10. $\frac{3}{12} + \frac{13}{12}$ işleminin sonucu kaçtır?</p> <p>a) $\frac{16}{24}$ b) $1 \frac{4}{12}$ c) $1 \frac{8}{12}$ d) $\frac{10}{12}$</p>	<p>Doğru cevaplayabilir</p> <p>()</p>	<p>Doğru cevaplayamaz</p> <p>()</p>
<p>Yüzde kaçını doğru cevaplayabilir?</p> <p>%</p>	<p>11. Ceyhan, sepetteki elmaların $\frac{3}{6}$'sını annesine, $\frac{2}{6}$'sını da arkadaşına veriyor. Geriye 24 tane elma kaldığına göre, başlangıçta sepette kaç tane elma vardı?</p> <p>a) 32 b) 36 c) 44 d) 144</p>	<p>Doğru cevaplayabilir</p> <p>()</p>	<p>Doğru cevaplayamaz</p> <p>()</p>

<p>Yüzde kaç doğru cevaplayabilir?</p> <p>%</p>	<p>12. Aşağıdaki şekilde bir bütün eş parçalarına bölünmüştür. Verilen kesirlerden hangisi taralı kısımlara karşılık gelir?</p>  <p>a) $\frac{4}{8}$ b) $\frac{4}{16}$ c) $\frac{5}{16}$ d) $\frac{6}{16}$</p>	<p>Doğru cevaplayabilir</p> <p>()</p>	<p>Doğru cevaplayamaz</p> <p>()</p>
<p>Yüzde kaç doğru cevaplayabilir?</p> <p>%</p>	<p>13. $323 \div 17$ işleminin sonucu kaçtır?</p> <p>a) 15 b) 17 c) 18 d) 19</p>	<p>Doğru cevaplayabilir</p> <p>()</p>	<p>Doğru cevaplayamaz</p> <p>()</p>

DEĞERLENDİRME 1	SORULAR	DEĞERLENDİRME 2	
<p>Yüzde kaçını doğru cevaplayabilir?</p> <p>%</p>	<p>14. $1\frac{2}{3}$ kesrinin sayı doğrusu üzerindeki okla gösterimi aşağıdakilerden hangisidir?</p> <p>a) </p> <p>b) </p> <p>c) </p> <p>d) </p>	<p>Doğru cevaplayabilir</p> <p>()</p>	<p>Doğru cevaplayamaz</p> <p>()</p>

DEĞERLENDİRME 1	SORULAR	DEĞERLENDİRME 2	
<p>Yüzde kaç doğru cevaplayabilir?</p> <p>%</p>	<p>18.</p>  <p>Yukarıdaki sayı doğrusunda A, B, C, D harfleri ile gösterilen noktalar aşağıdakilerden hangisinde doğru olarak verilmiştir?</p> <p>a) $A = \frac{4}{4}$; $B = \frac{2}{4}$; $C = \frac{8}{4}$; $D = \frac{2}{4}$</p> <p>b) $A = \frac{1}{4}$; $B = \frac{2}{4}$; $C = \frac{11}{4}$; $D = \frac{16}{4}$</p> <p>c) $A = \frac{4}{4}$; $B = \frac{9}{4}$; $C = \frac{11}{4}$; $D = \frac{14}{4}$</p> <p>d) $A = \frac{4}{4}$; $B = \frac{9}{4}$; $C = \frac{12}{4}$; $D = \frac{13}{4}$</p>	<p>Doğru cevaplayabilir</p> <p>()</p>	<p>Doğru cevaplayamaz</p> <p>()</p>

DEĞERLENDİRME 3

UYGUNLUK	ZORLUK		
	Kolay	Orta	Zor
Gerekli	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>
Önemli	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>
Kabul Edilebilir	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>
Tartışılabilir	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>	<i>Sorular:</i> <i>Yüzde:</i>

EK 3: Angoff Yöntemine Ait Kararlar

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
U1	60	45	40	80	75	60	30	40	90	80	40	60	60	50	60	40	30	20	70	60
U2	90	90	80	80	50	40	40	60	40	80	40	80	90	30	80	60	40	40	80	40
U3	70	50	60	90	40	60	40	40	70	90	60	90	70	60	70	80	60	60	50	30
U4	80	80	80	80	20	80	70	20	80	10	50	40	80	25	30	20	80	5	20	20
U5	70	60	50	80	70	50	5	40	50	70	5	50	70	10	90	80	20	10	60	20
U6	75	65	70	75	60	60	40	40	75	75	30	65	75	70	95	55	70	50	90	85
U7	70	70	40	30	40	70	25	30	50	40	50	35	70	15	33	40	60	15	80	70
U8	80	60	80	50	60	90	40	70	90	80	60	20	95	40	90	75	77	40	65	45
U9	25	25	40	60	25	25	5	10	50	50	10	25	25	25	25	10	5	10	25	20
U10	30	30	50	20	30	40	30	30	70	30	40	70	30	20	40	30	25	20	40	40
U11	60	30	70	30	40	40	20	40	50	25	30	80	60	60	30	60	30	60	70	60
U12	70	35	65	70	25	45	15	35	80	45	20	75	70	40	60	55	25	40	75	45
U13	65	40	50	70	60	55	45	60	90	60	45	90	45	50	50	65	40	50	60	70
U14	15	25	15	25	10	40	5	5	50	15	10	80	25	15	70	50	25	5	75	20
U15	40	60	50	45	60	80	10	30	70	45	20	100	45	45	100	85	75	35	100	95
U16	75	60	80	85	60	55	40	40	80	85	35	65	75	75	70	50	75	45	90	70
U17	60	60	60	70	40	45	50	40	60	70	50	50	70	30	55	55	60	20	70	50

EK 4: Yes/No Yöntemine Ait Kararlar

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
U1	1	0	0	1	1	1	0	0	1	1	0	1	1	0	1	0	0	0	1	1
U2	1	1	1	1	0	0	0	0	0	1	0	1	1	0	1	1	0	0	1	0
U3	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	0
U4	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	1	1	0	0	0
U5	1	0	0	1	1	0	0	0	0	1	0	0	1	0	1	1	0	0	1	0
U6	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	0	1	1	1	1
U7	1	1	1	0	0	1	0	0	1	0	1	0	1	0	0	1	1	0	1	1
U8	1	0	1	0	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1
U9	0	0	1	1	0	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0
U10	0	0	1	0	1	0	0	1	1	0	0	1	0	0	1	1	0	0	1	1
U11	1	0	1	0	1	1	0	1	1	0	0	1	1	1	1	1	0	0	1	1
U12	1	0	1	1	0	1	0	0	1	1	0	1	1	0	1	1	0	0	1	0
U13	1	0	0	1	1	0	0	1	1	1	0	1	0	0	0	1	0	0	0	1
U14	0	1	0	1	0	1	0	0	1	0	0	1	1	0	1	1	1	0	1	1
U15	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1	1	1	0	1	1
U16	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	0	1	0	1	1
U17	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1

EK 5: Madde Bazında Ebel Yöntemine Ait Kararlar

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
U1	65	35	40	75	75	55	35	35	75	80	35	55	65	55	60	40	35	35	65	55
U2	60	60	60	60	60	40	40	40	40	60	40	40	60	40	40	40	40	40	60	40
U3	60	60	60	90	40	50	50	40	60	90	40	90	60	60	60	80	60	60	60	40
U4	25	25	25	25	10	25	25	20	25	10	10	5	25	20	10	25	25	20	5	20
U5	60	50	50	60	60	50	10	40	50	60	10	60	50	10	60	60	10	10	60	10
U6	75	70	70	85	70	75	40	40	75	70	35	70	75	75	95	50	70	50	85	85
U7	70	70	50	35	30	70	25	30	35	40	50	60	70	35	50	35	60	15	60	50
U8	80	35	80	35	35	80	25	40	80	80	40	25	80	25	80	35	35	25	40	25
U9	25	25	40	40	20	10	10	20	50	50	15	40	25	20	15	15	5	10	20	5
U10	30	30	60	15	40	40	15	30	60	20	10	60	40	20	40	30	10	15	60	40
U11	50	15	50	60	60	40	15	40	60	15	15	60	50	60	60	50	40	60	60	60
U12	80	50	80	30	50	50	25	30	70	70	30	60	80	25	40	30	70	30	50	20
U13	60	45	60	65	65	45	50	75	75	70	45	75	45	50	50	75	45	60	50	60
U14	15	15	15	15	15	60	5	5	50	5	5	5	60	15	50	50	60	5	60	15
U15	75	75	45	45	60	65	10	45	80	45	20	75	65	45	80	80	65	35	75	80
U16	85	85	85	85	60	60	45	40	85	70	40	65	85	75	75	55	75	45	75	65
U17	70	60	65	70	50	60	50	50	70	60	65	60	65	30	55	60	65	30	60	60

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Ceylan Gündeğer

Doğum Yeri ve Tarihi : Ankara / 03.11.1986

Eğitim Durumu

Lisans Öğrenimi : Hacettepe Üniversitesi Sınıf Öğretmenliği ABD

Yüksek Lisans Öğrenimi : Eğitimde Ölçme ve Değerlendirme B.D.

Bildiği Yabancı Diller : İngilizce, Almanca

Bilimsel Faaliyetleri :

İş Deneyimi

Stajlar :

Projeler :

Çalıştığı Kurumlar : MEV Gökkuşuğu İlköğretim Okulu (2008-2009);
Aksaray Üniversitesi - Hacettepe Üniversitesi
(2010-...)

İletişim

E-Posta Adresi : cgundeger@gmail.com

Tarih : 13.01.2012